

Result No.	Query			DB	ID	Description	
	Score	Match	Length				
1	589.4	99.7	591	9	AY408336	Homo sapi	
2	589.4	99.7	729	7	CN304585	170006005	
3	589.4	99.7	742	7	CN304584	170005979	
4	589.4	99.7	1038	4	BM560308	AGENCOURT	
5	589.4	99.7	1298	3	CR605181	full-length	
6	589.4	99.7	1318	3	CR601245	full-length	
7	582.6	98.6	1056	5	BU157342	AGENCOURT	
8	580.4	98.2	646	7	CN304590	170006001	
9	575.8	97.4	1017	5	BX398356	EX398356	
10	575.2	97.3	712	7	CN304594	170005326	
11	567.8	96.1	870	5	BU508982	AGENCOURT	
12	560.8	94.9	973	7	CO579535	ILLUMIGEN	
13	559.8	94.7	668	4	BG720189	602692233	
14	558	94.4	1001	5	BX362283	EX362283	
15	557	94.2	1045	4	BM546306	AGENCOURT	
16	529	89.5	612	5	EX113979	EX113979	
17	528.8	89.5	649	4	BM716519	UI-E-EJO-	
18	527	89.2	575	7	CV028541	7081_Full1	
19	523.4	88.6	642	4	BG105954	602290144	
20	517.8	87.6	649	7	CN304591	170004245	
21	501	84.8	838	6	CB995481	AGENCOURT	
22	499.6	84.5	705	4	BG527376	602557362	
23	494.4	83.7	598	7	CN304593	170005977	
24	482.4	81.6	568	5	BP262533	BP262533	

	Query Match	99.7%	Score 589.4	DB 9	Length 591
	Best Local Similarity	99.8%	Pred. No. 1.6e-135		
	Matches 590	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Qy	1	ATGGAGGGGAGGCTTGAAGAGGAAACA	CTCTGATTTTCGAAGAGGAGGAGGAGGTGGGAG	60	
Db	1	ATGGTGTGGAGGCTTGAAGAGGAAACA	CTCTGATTTTCGAAGAGGAGGAGGAGGTGGGAG	60	

QY 61 TGGAGTCCAGCAGGCTTTCAGAGCTACCAAGCCCTGCTCGCATCTCCCTAGACAAA 120
DB 61 TGGAGTCCAGCAGGCTTTCAGAGCTACCAAGCCCTGCTCGCATCTCCCTAGACAAA 120
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCCAGCTCCGAGGAGTGTCTCATCATCAAC 180
DB 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCCAGCTCCGAGGAGTGTCTCATCATCAAC 180
QY 181 ACCCTCCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 240
DB 181 ACCCTCCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 240
QY 241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
DB 241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
QY 301 GAGCTGGACACCTCCATGGATGGGAGTGGAGCTGAGCCCTCAGAACTCCAGTGACTCCCTTTGGC 360
DB 301 GAGCTGGACACCTCCATGGATGGGAGTGGAGCTGAGCCCTCAGAACTCCAGTGACTCCCTTTGGC 360
QY 361 CTCAGAGTGAAGTGCACCCAGCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 420
DB 361 CTCAGAGTGAAGTGCACCCAGCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 420
QY 421 TACTTTGGGAGCTCTGGCTGATGACTTCTTCTTGGACATTCACACATCTCGGCTAGAA 480
DB 421 TACTTTGGGAGCTCTGGCTGATGACTTCTTCTTGGACATTCACACATCTCGGCTAGAA 480
QY 481 AAGGAGCTTCGACGGGCCCCCAGAGGCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 540
DB 481 AAGGAGCTTCGACGGGCCCCCAGAGGCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 540
QY 541 TGGGAGTGAATGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
DB 541 TGGGAGTGAATGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591

RESULT 2
CN304585
LOCUS 17000597912783 GRN_PRENEP Homo sapiens cDNA 5', mRNA linear EST 16-MAY-2004
DEFINITION 17000597912783 GRN_PRENEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN304585
VERSION CN304585.1 GI:47320999
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 729)
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,
Lebkowski, J. and Stanton, L. W.
TITLE Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: xbrandenberger@geron.com
Insert Length: 729 Std Error: 0.00.
Location/Qualifiers
1. .729
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cell, retinoic acid and mitogen-treated hES cell line H7"
/clone_lib="GRN PRENEP"
/note="oligo dt primed, full-length enriched cDNA library"

FEATURES
source

ORIGIN

Query Match 99.7%; Score 589.4; DB 7; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-135;
Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGAGGAGGCTTCAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 60
DB 98 ATGAGGAGGCTTCAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 157
QY 61 TGGAGTCCAGCAGGCTTTCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 120
DB 158 TGGAGTCCAGCAGGCTTTCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 217
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCCAGCTCCGAGGAGTGTCTCATCATCAAC 180
DB 218 GTCCAGCGCAGCTGGGCCCCCGAGCACCCAGCTCCGAGGAGTGTCTCATCATCAAC 277
QY 181 ACCCTCCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 240
DB 278 ACCCTCCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 337
QY 241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
DB 338 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 397
QY 301 GAGCTGGACACCTCCATGGATGGGAGTGGAGCTGAGCCCTCAGAACTCCAGTGACTCCCTTTGGC 360
DB 398 GAGCTGGACACCTCCATGGATGGGAGTGGAGCTGAGCCCTCAGAACTCCAGTGACTCCCTTTGGC 457
QY 361 CTCAGAGTGAAGTGCACCCAGCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 420
DB 458 CTCAGAGTGAAGTGCACCCAGCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 517
QY 421 TACTTTGGGAGCTCTGGCTGATGACTTCTTCTTGGACATTCACACATCTCGGCTAGAA 480
DB 518 TACTTTGGGAGCTCTGGCTGATGACTTCTTCTTGGACATTCACACATCTCGGCTAGAA 577
QY 481 AAGGAGCTTCGACGGGCCCCCAGAGGCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 540
DB 578 AAGGAGCTTCGACGGGCCCCCAGAGGCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 637
QY 541 TGGGAGTGAATGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
DB 638 TGGGAGTGAATGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 688

RESULT 3
CN304584

LOCUS 1700059791761 GRN_PREHEP Homo sapiens cDNA 5', mRNA linear EST 16-MAY-2004
DEFINITION 1700059791761 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN304584
VERSION CN304584.1 GI:47320998
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 742)
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,
Lebkowski, J. and Stanton, L. W.
TITLE Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658

Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 742 Std Error: 0.00.
Location/Qualifiers
1. .742
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, DMSO-treated H9 cell line"
/clone_lib="GRN PREHEP"
/note="oligo dt primed, full-length enriched cDNA library from DMSO-treated hES cell line H9 (p22) maintained in feeder-free conditions"

REFERENCE
1 (bases 1 to 1038)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMW12766 row: i column: 08
High quality sequence start: 32
High quality sequence stop: 703.

FEATURES
source
1. 1038
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5744719"
/tissue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 99.7%; Score 589.4; DB 7; Length 742;
Best Local Similarity 99.8%; Pred. No. 1.7e-135;
Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGAGGGAGGCTTGAAGAGGAACAACCTCTGATTTGGAAAGAGGAGGAGGTGGGAG 60
DB 68 ATGTTGGGAGGCTTGAAGAGGAACAACCTCTGATTTGGAAAGAGGAGGAGGTGGGAG 127
QY 61 TGGAGTCCAGCAGGCTTCAGAGCTACAGAGCCCTGCTCGCATCTCCCTAGACAAA 120
DB 128 TGGAGTCCAGCAGGCTTCAGAGCTACAGAGCCCTGCTCGCATCTCCCTAGACAAA 187
QY 121 GTCCAGCGAGCTGGGCCCCCGAGCACCAGCTCCGAGCAGCATGCTCTCATCCATAAC 180
DB 188 GTCCAGCGAGCTGGGCCCCCGAGCACCAGCTCCGAGCAGCATGCTCTCATCCATAAC 247
QY 181 ACCCTCCAAACAGCTGCAGGCTGCACATTCGCTGGCTCCGCCCCCTGCTCGCATCTCCCTAG 240
DB 248 ACCCTCCAAACAGCTGCAGGCTGCACATTCGCTGGCTCCGCCCCCTGCTCGCATCTCCCTAG 307
QY 241 CCCCTCTTCTGGGCGAGAGATTTCTCCCTGTGACCCACCATTTGGCTCTATCTCAGG 300
DB 308 CCCCTCTTCTGGGCGAGAGATTTCTCCCTGTGACCCACCATTTGGCTCTATCTCAGG 367
QY 301 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 360
DB 368 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 427
QY 361 CTCAGAGTGAAGTGCACCCAGCCTGATCCAGTCTTCTTGAAGAGCTCTGAGCTCCCGG 420
DB 428 CTCAGAGTGAAGTGCACCCAGCCTGATCCAGTCTTCTTGAAGAGCTCTGAGCTCCCGG 487
QY 421 TACTTGGGGAGCTCTGGCTGGAGTGAATTTCTTTCTGGAACATTTGACACATCTGGGTAGAA 480
DB 488 TACTTGGGGAGCTCTGGCTGGAGTGAATTTCTTTCTGGAACATTTGACACATCTGGGTAGAA 547
QY 481 AAGGAGCTGCACGGGCCCCCAGAGCTCTCACAACCTTCTGTGCCCCCAGGTTCT 540
DB 548 AAGGAGCTGCACGGGCCCCCAGAGCTCTCACAACCTTCTGTGCCCCCAGGTTCT 607
QY 541 TGGGAGTGAAGTGAATTCGATCAGATGGAATTCATTTCTGGGTCTTAA 591
DB 608 TGGGAGTGAAGTGAATTCGATCAGATGGAATTCATTTCTGGGTCTTAA 658

RESULT 4
BM560308
LOCUS
DEFINITION
AGENCOURT_6563953 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5744719
5', mRNA sequence.
ACCESSION
BM560308
VERSION
BM560308.1 GI:18804629
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

|||||
Db 580 AAGGAGCCTGCACGGGCCCCCAGAGCCTCCTCAACACTCTCTGTGCCCCAGGTTCT 639
|||||
QY 541 TGGGAGTGAAGTAAGTGGATCAGATCAGATCAGGAAATCATTCTGGGTCCTAA 591
|||||
Db 640 TGGGAGTGAAGTAAGTGGATCAGATCAGATCAGGAAATCATTCTGGGTCCTAA 690
|||||

RESULT 5
LOCUS CR605181 1298 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DJ012YJ05 of T cells (Jurkat cell line)
ACCESSION CR605181
VERSION CR605181.1 GI:50485988
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1298)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1298)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..1298
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ012YJ05"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 99.7%; Score 589.4; DB 3; Length 1298;
Best Local Similarity 99.8%; Pred. No. 1.9e-135;
Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTGGAGAGGAGGAGAGGTGGGAG 60
Db 105 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTGGAGAGGAGGAGAGGTGGGAG 164
QY 61 TGGAGTCCAGCAGGCTTTCAGAGCTACCGAAGCCCTGCTCCGATCTCCCTAGACAA 120
Db 165 TGGAGTCCAGCAGGCTTTCAGAGCTACCGAAGCCCTGCTCCGATCTCCCTAGACAA 224
QY 121 GTCCAGCGCAGCTTGGGCCCCGAGCAGCCGCTCCGAGCAGTGTCTCATCCATAC 180
Db 225 GTCCAGCGCAGCTTGGGCCCCGAGCAGCCGCTCCGAGCAGTGTCTCATCCATAC 284
QY 181 ACCCTCCAAACAGCTGCAGGCTGACATTGCGCTGGCTCCGCCCCCTCGCCCTGCCCCCGAG 240
Db 285 ACCCTCCAAACAGCTGCAGGCTGACATTGCGCTGGCTCCGCCCCCTCGCCCTGCCCCCGAG 344
QY 241 CCCCTCTTCTGGGCGAGGAGATTTCCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
Db 345 CCCCTCTTCTGGGCGAGGAGATTTCCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 404

```

```

QY 301 GAGCTGGACACCTCCATCGATGGGAGCTGAGCCCCCTCAGAATCCAGTGAAGTCCCTTGGC 360
Db 405 GAGCTGGACACCTCCATCGATGGGAGCTGAGCCCCCTCAGAATCCAGTGAAGTCCCTTGGC 464
QY 361 CTCAGAATGAAGTGGCCACCCAGGCTGATCAGTCTTCTTAGAAGCTCTGAGTCCCGG 420
Db 465 CTCAGAATGAAGTGGCCACCCAGGCTGATCAGTCTTCTTAGAAGCTCTGAGTCCCGG 524
QY 421 TACTTTGGGGGACTCTGGCTCGATGAGTCTTTCTTGGACATTTGACACATCTGCGGTAGAA 480
Db 525 TACTTTGGGGGACTCTGGCTCGATGAGTCTTTCTTGGACATTTGACACATCTGCGGTAGAA 584
QY 481 AAGGAGCCTGCACGGGCCCCCAGAGCCTCCTCAACACCTCTTCTGTGCCCCAGGTTCT 540
Db 585 AAGGAGCCTGCACGGGCCCCCAGAGCCTCCTCAACACCTCTTCTGTGCCCCAGGTTCT 644
QY 541 TGGGAGTGAATGAAGTGAAGTGAATGATCATATGAAATCATTTCTGGGGTCTCTAA 591
Db 645 TGGGAGTGAATGAAGTGAAGTGAATGATCATATGAAATCATTTCTGGGGTCTCTAA 695

RESULT 6
LOCUS CR601245 1318 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DI057YM14 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR601245
VERSION CR601245.1 GI:50482052
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1318)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1318)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..1318
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI057YM14"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 99.7%; Score 589.4; DB 3; Length 1318;
Best Local Similarity 99.8%; Pred. No. 1.9e-135;
Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTGGAGAGGAGGAGAGGTGGGAG 60
Db 112 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTGGAGAGGAGGAGAGGTGGGAG 171
QY 61 TGGAGTCCAGCAGGCTTTCAGAGCTACCGAAGCCCTGCTCCGATCTCCCTAGACAA 120
Db 172 TGGAGTCCAGCAGGCTTTCAGAGCTACCGAAGCCCTGCTCCGATCTCCCTAGACAA 231

```

QY 121 GTCCAGCAGCCTGGGCCCCCGAGCACCAGCCTCCGCGAGCATGTCTCATCCATAAC 180
 Db 232 GTCCAGCAGCCTGGGCCCCCGAGCACCAGCCTCCGCGAGCATGTCTCATCCATAAC 291
 QY 181 ACCCTCCAAACAGCTGCAGGCTGCACCTTGGCTGGCTCCGCGCCCTGCGCTGCCCGCCGAG 240
 Db 292 ACCCTCCAAACAGCTGCAGGCTGCACCTTGGCTGGCTCCGCGCCCTGCGCTGCCCGCCGAG 351
 QY 241 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGTGAGCCACCATTTGGCTCTATCTCAGG 300
 Db 352 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGTGAGCCACCATTTGGCTCTATCTCAGG 411
 QY 301 GAGCTGGACACCTCCATGATGGGATGGGACTGAGCCCTCAGAACTCCAGTGAATCTCCCTTGGC 360
 Db 412 GAGCTGGACACCTCCATGATGGGATGGGACTGAGCCCTCAGAACTCCAGTGAATCTCCCTTGGC 471
 QY 361 CTCAGAGTGAAGTGCACCCAGCCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 420
 Db 472 CTCAGAGTGAAGTGCACCCAGCCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 531
 QY 421 TACTTGGGGGACTCTGGGCTGATGATCTTTCTTGGACATTTGACATCTCTCGGTAGAA 480
 Db 532 TACTTGGGGGACTCTGGGCTGATGATCTTTCTTGGACATTTGACATCTCTCGGTAGAA 591
 QY 481 AAGGAGCTGACGCGGCCCCCAGAGCCTCTCACAACCTCTTCTGTGCCCGAGTTCT 540
 Db 592 AAGGAGCTGACGCGGCCCCCAGAGCCTCTCACAACCTCTTCTGTGCCCGAGTTCT 651
 QY 541 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
 Db 652 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 702

RESULT 7
 BUL57342 1056 bp mRNA linear EST 04-SEP-2002
 Locus AGENCOURT_6925371 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5952523
 5', mRNA sequence.

ACCESSION BUL57342
 VERSION BUL57342.1 GI:22670874
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue procurement: ATCC

CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCW2140 row: k column: 20
 High quality sequence stop: 587.
 Location/Qualifiers

FEATURES
 source
 1..1056
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5952523"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 110"
 /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
 Site 2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by

ORIGIN

Query Match 98.6%; Score 582.6; DB 5; Length 1056;
 Best Local Similarity 99.0%; Pred. No. 8.8e-134;
 Matches 585; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 ATGAGGAGGAGGCTTCAAGAGGAACACCTCTGATTTGGAGAGGAGGAGGAGGAGGAGGAG 60
 Db 108 ATGAGGAGGAGGCTTCAAGAGGAACACCTCTGATTTGGAGAGGAGGAGGAGGAGGAGGAG 167
 QY 61 TGGAGTCCAGCAGGCGCTTTCAGAGCTTACAGCAGAGCCCTGCTCCGATCTTCCCTAGACAAA 120
 Db 168 TGGAGTCCAGCAGGCGCTTTCAGAGCTTACAGCAGAGCCCTGCTCCGATCTTCCCTAGACAAA 227
 QY 121 GTCCAGCAGCCTGGGCCCCCGAGCACCAGCCTCCGCGAGCATGTCTCATCCATAAC 180
 Db 228 GTCCAGCAGCCTGGGCCCCCGAGCACCAGCCTCCGCGAGCATGTCTCATCCATAAC 287
 QY 181 ACCCTCCAAACAGCTGCAGGCTGCACCTTGGCTGGCTCCGCGCCCTGCGCTGCCCGCCGAG 240
 Db 288 ACCCTCCAAACAGCTGCAGGCTGCACCTTGGCTGGCTCCGCGCCCTGCGCTGCCCGCCGAG 347
 QY 241 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
 Db 348 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 407
 QY 301 GAGCTGGACACCTCCATGATGGGACTGAGCCCTCAGAACTCCAGTGAATCTCCCTTGGC 360
 Db 408 GAGCTGGACACCTCCATGATGGGACTGAGCCCTCAGAACTCCAGTGAATCTCCCTTGGC 467
 QY 361 CTCAGAGTGAAGTGCACCCAGCCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 420
 Db 468 CTCAGAGTGAAGTGCACCCAGCCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 527
 QY 421 TACTTGGGGGACTCTGGGCTGATGATCTTTCTTGGACATTTGACACATCTCGGCTAGAA 480
 Db 528 TACTTGGGGGACTCTGGGCTGATGATCTTTCTTGGACATTTGACACATCTCGGCTAGAA 587
 QY 481 AAGGAGCTGACGCGGCCCCCAGAGCCTCTCACAACCTCTTCTGTGCCCGAGGTTCT 540
 Db 588 AAGGAGCTGACGCGGCCCCCAGAGCCTCTCACAACCTCTTCTGTGCCCGAGGTTCT 647
 QY 541 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
 Db 648 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 698

RESULT 8

CN304590 646 bp mRNA linear EST 16-MAY-2004
 Locus 17000600179006 GRN_PNEU Homo sapiens cDNA 5', mRNA sequence.

DEFINITION CN304590

ACCESSION CN304590.1 GI:47321004

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 646)

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,

Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,

Lebkowski, J. and Stanton, L.W.

Transcriptome characterization elucidates signaling networks that

control human ES cell growth and differentiation

Nat. Biotechnol. 22 (6), 707-716 (2004)

Contact: Brandenberger R

Regenerative Medicine

Geron Corporation

230 Constitution Drive, Menlo Park, CA 94025, USA

Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."


```

QY 541 TGGGAGTGAATGAATGGATCAGATCAGATCGAATCATTTCTGGGCTCTTAA 591
Db 651 TGGGAGTGAATGAATGGATCAGATCAGATCGAATCATTTCTGGGCTCTTAA 701

RESULT 10
LOCUS CN304594 712 bp mRNA linear EST 16-MAY-2004
DEFINITION L7000532611443 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN304594
VERSION CN304594.1 GI:47321008
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 712)
AUTHORS Li Y., Xu C., Fang R., Guegler K., Rao M.S., Mandalam R.,
Lebkowski J. and Stanton L.W.
TITLE Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 712 Std Error: 0.00.
Location/Qualifiers
FEATURES
source
1..712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, embryoid bodies
derived from HL, H7 and H9 cells"
/clone_lib="GRN EB"
/note="oligo dt primed, full-length enriched cDNA library
from embryoid body outgrowths derived from hES cell lines
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
conditions."

ORIGIN
Query Match 97.3%; Score 575.2; DB 7; Length 712;
Best Local Similarity 99.3%; Pred. No. 5.6e-132;
Matches 588; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATGGAGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGAG 60
Db 115 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGAG 174
QY 61 TGGAGTCCAGAGGCTTCAAGAGTACCAGAGCCCTCTCCGATCTCCCTAGACAAA 120
Db 175 TGGAGTCCAGAGGCTTCAAGAGTACCAGAGCCCTCTCCGATCTCCCTAGACAAA 234
QY 121 GTCCAGCGCAGCTGGGCCCCGAGCAGCCAGCTCCGAGGATGTCTCATCATTAAC 180
Db 235 GTCCAGCGCAGCTGGGCCCCGAGCAGCCAGCTCCGAGGATGTCTCATCATTAAC 294
QY 181 ACCCTCCAGAGCTGAGGCTGCACTTGGCTGGCTCCGCCCCCTGCCCCCGAG 240
Db 295 ACCCTCCAGAGCTGAGGCTGCACTTGGCTGGCTCCGCCCCCTGCCCCCGAG 354
QY 241 CCCCTCTTCTGGGCGAGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCCTCAGG 300
Db 355 CCCCTCTTCTGGGCGAGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCCTCAGG 414
QY 301 GAGCTGGACACCTCCATGATGGGACTGAGCCCCCTCAGAAATCCAGTGAATCTCCCTTGGC 360
Db 415 GAGCTGGACACCTCCATGATGGGACTGAGCCCCCTCAGAAATCCAGTGAATCTCCCTTGGC 474

```

```

QY 361 CTCAGAGTGAAGTCCACCCAGCCTGATCAGTCTTTCTAGAGCTCTGAGCTCCCGG 420
Db 475 CTCAGAGTGAAGTCCACCCAGCCTGATCAGTCTTTCTAGAGCTCTGAGCTCCCGG 534
QY 421 TACTTGGGGGACTCTGGGCTGGATGACTTCTTTCTGGACATTTGACACATCTCGCGTAGAA 480
Db 535 TACTTGGGGGACTCTGGGCTGGATGACTTCTTTCTGGACATTTGACACATCTCGCGTAGAA 594
QY 481 AAGGAGCTTGCACGGGCCCCACCGAGAGCTCTCTCAACCTCTTCTGTGCCCCAGGTT-C 539
Db 595 AAGGAGCTTGCACGGGCCCCACCGAGAGCTCTCTCAACCTCTTCTGTGCCCCAGGTTCC 654
QY 540 TTGGGAGTGAATGAATCGGATCAGATCAGTGAATCATTTCTGGGCTCTTAA 591
Db 655 TGGGAGTGAATGAATCGGATCAGATCAGTGAATCATTTCTGGGCTCTTAA 706

RESULT 11
LOCUS BUS08982 873 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT_10095430 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6503515
5', mRNA sequence.
ACCESSION BUS08982
VERSION BUS08982.1 GI:22815215
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 873)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM14061 row: e column: 20
High quality sequence stop: 605.
Location/Qualifiers
FEATURES
source
1..873
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:6503515"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

ORIGIN
Query Match 96.1%; Score 567.8; DB 5; Length 873;
Best Local Similarity 98.5%; Pred. No. 4e-130;
Matches 583; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1 ATGGAGGAGGCTTCAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGAG 60
Db 179 ATGGTGGGAGGCTTCAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGAG 238
QY 61 TGGAGTCCAGAGGCTTCAAGAGTACCAGAGCCCTGCTCCGATCTCCCTAGACAAA 120
Db 239 TGGAGTCCAGAGGCTTCAAGAGTACCAGAGCCCTGCTCCGATCTCCCTAGACAAA 298
QY 121 GTCCAGCGCAGCTGGGCCCCGAGCAGCCAGCTCCGAGGATGTCTCATCATTAAC 180
Db 299 GTCCAGCGCAGCTGGGCCCCGAGCAGCCAGCTCCGAGGATGTCTCATCATTAAC 358

```


Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10736 row: a column: 09
High quality sequence stop: 666.
Location/Qualifiers
1. .668
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4824464"
/lab_host="DH10B"
/clone_lib="NIH MGC 97"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
(gtcag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.2 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this is
a NIH MGC Library."

ORIGIN

Query Match 94.7%; Score 559.8; DB 4; Length 668;
Best Local Similarity 99.5%; Pred. No. 3.7e-128;
Matches 561; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGGAG 60
DB 105 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGGAG 164
QY 61 TGGAGTCAGAGGCGCTTACAGAGCTACAGAGCGCTCTCGCATCTCCCTACATAC 120
DB 165 TGGAGTCCAGAGGCGCTTACAGAGCTACAGAGCGCTCTCGCATCTCCCTACATAC 224
QY 121 GTCCAGGCGAGCTGGGCCCCGAGCAGCCAGCTCCGAGGAGTGTCTCATCATAC 180
DB 225 GTCCAGGCGAGCTGGGCCCCGAGCAGCCAGCTCCGAGGAGTGTCTCATCATAC 284
QY 181 ACCCTCCAAAGCTGAGGCTGACATTTCCCTGTGAGTCCGCGCTGCTCCGCTGCTCCGCGAG 240
DB 285 ACCCTCCAAAGCTGAGGCTGACATTTCCCTGTGAGTCCGCGCTGCTCCGCTGCTCCGCGAG 344
QY 241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGTCCGCGCTGCTCCGCTGCTCCGCGAG 300
DB 345 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGTCCGCGCTGCTCCGCTGCTCCGCGAG 404
QY 301 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 360
DB 405 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 464
QY 361 CTCAGAGTGAAGTGCACCCAGGCTGATCCAGTCTTCTTGAAGAGCTCTGAGCTCCCGG 420
DB 465 CTCAGAGTGAAGTGCACCCAGGCTGATCCAGTCTTCTTGAAGAGCTCTGAGCTCCCGG 524
QY 421 TACTTGGGGGAGCTCTGGGCTGGAGTGAATCTTCTTCTTGGACATTTGACACATCTCGGTAGAA 480
DB 525 TACTTGGGGGAGCTCTGGGCTGGAGTGAATCTTCTTCTTGGACATTTGACACATCTCGGTAGAA 584
QY 481 AAGGAGCTGACAGGCCCCCAGCAGGCTCTCTCAACCTCTTCTGTGCCCCCAGGTTCT 540
DB 585 AAGGAGCTGACAGGCCCCCAGCAGGCTCTCTCAACCTCTTCTGTGCCCCCAGGTTCT 644
QY 541 TGGGAGTGGAAATGAATGGATCAC 564
DB 645 TGGGAGTGGAAATGAATGGATCAC 668

RESULT 14
BX362283
LOCUS
DEFINITION
BX362283 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CSODJ012YJ05 5-PRIME, mRNA sequence.
ACCESSION
BX362283
VERSION
BX362283.2 GI:46289607
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 1001)
AUTHORS
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
On May 5, 2003 this sequence version replaced gi:30370588.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
9131.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CSODJ012CE03QP1&c=9131.r.
Location/Qualifiers
1..1001
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODJ012YJ05"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 94.4%; Score 558; DB 5; Length 1001;
Best Local Similarity 99.1%; Pred. No. 1.1e-127;
Matches 581; Conservative 1; Mismatches 2; Indels 2; Gaps 2;

QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGGAG 60
DB 105 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGTGGGAG 164
QY 61 TGGAGTCCAGAGGCGCTTACAGAGCTACAGAGCGCTTCCGCTGCTCCGCTGCTCCCTACATAC 120
DB 165 TGGAGTCCAGAGGCGCTTACAGAGCTACAGAGCGCTTCCGCTGCTCCGCTGCTCCCTACATAC 224
QY 121 GTCCAGGCGAGCTGGGCCCCGAGCAGCCAGCTCCGAGGAGTGTCTCATCATAC 180
DB 225 GTCCAGGCGCA -CTTGGGCCCCGAGCAGCCAGCTCCGAGGAGTGTCTCATCATAC 283
QY 181 ACCCTCCAAAGCTGAGGCTGACATTTCCCTGTGAGTCCGCGCTGCTCCGCTGCTCCGCGAG 240
DB 284 ACCCTCCAAAGCTGAGGCTGACATTTCCCTGTGAGTCCGCGCTGCTCCGCTGCTCCGCGAG 343
QY 241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGTCCGCGCTGCTCCCTGCTCATCTCAGG 300
DB 344 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGTCCGCGCTGCTCCCTGCTCATCTCAGG 403
QY 301 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 360
DB 404 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 463

```
QY 361 CTCAGAGTGAAGTCCACCCAGCCTGATCCAGTCTTTCTTAGAAGCTCTGAGCTCCCGG 420
Db 464 CTCAGAGTGAAGTCCACCCAGCCTGATCCAGTCTTTCTTAGAAGCTCTGAGCTCCCGG 523
QY 421 TACTTTGGGGGACTCTGGCCTGTGATGACTTCTTTCTGGACATTTGACACATCTCGGGTAGAA 480
Db 524 TACTTTGGGGGACTCTGGCCTGTGATGACTTCTTTCTGGACATTTGACACATCTCGGGTAGAA 583
QY 481 AAGGAGCTGTCACGGGCCCCACAGAGCCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 540
Db 584 AAGGAGCTGTCACGGGCCCCACAGAGCCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 643
QY 541 TG-GGAGTGGGAATGAATGGATGACATCATGATGAAATCATTTCTGGGG 585
Db 644 TCGGAGTGGGAATGAATGGATGACATCATGATGAAATCATTTCTCGGG 689

RESULT 15
BM546306
LOCUS BM546306
DEFINITION AGENCOURT_6498625 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5722856
5', mRNA sequence.
ACCESSION BM546306
VERSION BM546306.1 GI:18779156
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1045)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12709 row: j column: 09
High quality sequence start: 20
High quality sequence stop: 626.
FEATURES
Location/Qualifiers
1..1045
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5722856"
/lab_host="DH10B"
/clone_lib="NIH_MGC_125"
/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."

ORIGIN
Query Match 94.2%; Score 557; DB 4; Length 1045;
Best Local Similarity 98.6%; Pred. No. 2e-127;
Matches 582; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 1 ATGAGGGGGGCTTCAAGAGGAACACTCTGATTTGGAGAGGAGGAGGTTGGGAG 60
Db 208 ATGGTGGGAGGGCTTGAAGAGGAACACTCTGATTTGGAGAGGAGGAGGTTGGGAG 267
QY 61 TGGAGTCCAGCAGGCGCTTTCAGAGCTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 120
```

```
Db 268 TGGAGTCCAGCAGGCGCTTTCAGAGCTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 327
QY 121 GTCCAGCGCAGCCTGGGCCCCGAGCACCAGCCTCCGCGAGCATGTCTCATCCATTAAC 180
Db 328 GTCCAGCGCAGCCTGGGCCCCGAGCACCAGCCTCCGCGAGCATGTCTCATCCATTAAC 387
QY 181 ACCCTCCAAAGCTGTCAGGCTGCACTTCGCTGGGCTCCCGCCCCCTGCGCTGCCCCCGAG 240
Db 388 ACCCTCCAAAGCTGTCAGGCTGCACTTCGCTGGGCTCCCGCCCCCTGCGCTGCCCCCGAG 447
QY 241 CCCCTTTCTCTGGGCGAGGAGGATTTCTCCCTGTGTCAGCCACCATTTGGTCTTATCTCAGG 300
Db 448 CCCCTTTCTCTGGGCGAGGAGGATTTCTCCCTGTGTCAGCCACCATTTGGTCTTATCTCAGG 507
QY 301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTCACTCCCCCTGGC 360
Db 508 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTCACTCCCCCTGGC 567
QY 361 CTCAGAGTGAAGTGCACCCAGCCTGATCCAGTCTTTCTTAGAAGCTCTGAGCTCCCGG 420
Db 568 CTCAGAGTGAAGTGCACCCAGCCTGATCCAGTCTTTCTTAGAAGCTCTGAGCTCCCGG 627
QY 421 TACTTTGGGGGACTCTGGCCTGTGATGACTTTCTTTCTGGACATTTGACACATCTCGGTAGAA 480
Db 628 TACTTTGGGGGACTCTGGCCTGTGATGACTTTCTTTCTGGACATTTGACACATCTCGGTAGAA 687
QY 481 AAGGAGCTGTCACGGGCCCCACAGAGCCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 540
Db 688 AAGGAGCTGTCACGGGCCCCACAGAGCCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 747
QY 541 TGGGAGTGGG-AATGAACTGGATCAGATCATCAT-GGAAATCATTTCTGGGGTCC 588
Db 748 TGGGAGTGGGCAATGACCTGGATCCCTCATGGGAAATCATTTCTGGGGTCC 797
```

Search completed: April 26, 2005, 10:11:12
Job time : 3125 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2005, 12:21:55 ; Search time 166 Seconds
(without alignments)
456.657 Million cell updates/sec

Title: US-10-069-386A-2
Perfect score: 1033
Sequence: 1 MEGGLKRXHSLDEEERWE.....APGSWNEHLDHIMEILIGS 196

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:.*
1: Geneseqp1980a:.*
2: Geneseqp1990a:.*
3: Geneseqp2000a:.*
4: Geneseqp2001a:.*
5: Geneseqp2002a:.*
6: Geneseqp2003a:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004a:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1033	100.0	196	4	AAB35402	Aab35402 Replicati
2	1026	99.3	196	4	AAM93922	Aam93922 Human pol
3	1026	99.3	196	8	ADL32052	Adl32052 Human pro
4	791	76.6	157	4	AAG74742	Aag74742 Human col
5	645.5	62.5	142	8	ABO60075	AbO60075 Human gen
6	402	38.9	111	8	ABO60074	AbO60074 Human gen
7	216	20.9	236	2	AAY36004	Aay36004 Extended
8	216	20.9	236	8	ADP19312	Adp19312 Human sec
9	213	20.6	236	3	AAV44362	Aav44362 Human cel
10	213	20.6	236	4	AAM93724	Aam93724 Human pol
11	213	20.6	236	4	AAG89292	Aag89292 Human sec
12	213	20.6	236	8	ADL31644	Adl31644 Human pro
13	213	20.6	278	4	AAM25550	Aam25550 Human pro
14	211	20.4	236	2	AAV31829	Aav31829 Human adu
15	205	19.8	222	8	ABO60367	AbO60367 Human gen
16	205	19.8	236	2	AAV02619	Aav02619 Amino aci
17	136	13.2	237	5	AAO22897	Aao22897 Mouse hae
18	125	12.1	314	5	ABP65076	Abp65076 Hypoxia-1
19	125	12.1	314	8	ADR14626	Adr14626 Human NF-
20	106	10.3	241	4	AAM39725	Aam39725 Human pol
21	106	10.3	254	3	AAB58258	Aab58258 Lung canc
22	106	10.3	254	4	AAG73682	Aag73682 Human col
23	106	10.3	254	4	AAM41511	Aam41511 Human pol
24	105.5	10.2	578	4	AAM38707	Aam38707 Human pol
25	105.5	10.2	620	7	ADN95361	Adn95361 Human BRC

26	105.5	10.2	620	8	ADP54705	Adp54705 Human PRO
27	105.5	10.2	650	4	AAM38706	Aam38706 Human pol
28	105.5	10.2	685	7	ADE28201	AdE28201 Human MDP
29	105.5	10.2	759	4	AAM40492	Aam40492 Human pol
30	105.5	10.2	759	4	AAM40493	Aam40493 Human pol
31	105.5	10.2	760	7	ADN95128	Adn95128 Human LEC
32	105	10.2	241	5	ABO6375	AbO6375 Human CHD
33	105	10.2	241	5	ABO22898	AbO22898 Human hae
34	105	10.2	241	8	ABM81593	Abm81593 Tumour-as
35	104.5	10.1	740	4	ABB11713	Abb11713 Human KIA
36	104.5	10.1	879	5	ABP69285	Abp69285 Human pol
37	104	10.1	322	4	ABG22868	Abg22868 Novel hum
38	101.5	9.8	728	7	ADD46841	AdD46841 Human PRO
39	101.5	9.8	728	7	ADJ66562	AdJ66562 Human PRO
40	101.5	9.8	728	8	ADJ66562	AdJ66562 P13 Kinas
41	98	9.5	562	7	ADR65060	Adr65060 Human PRO
42	98	9.5	562	8	ADR58973	Adr58973 Human Elk
43	98	9.5	562	8	ADR58971	Adr58971 Human Elk
44	97	9.4	803	8	ADN19883	Adn19883 Bacterial
45	96	9.3	817	6	ABR53281	AbR53281 Protein s

ALIGNMENTS

RESULT 1
AAB35402
ID AAB35402 standard; protein; 196 AA.
XX
AC AAB35402;
XX
DT 23-MAY-2001 (first entry)
XX
DE Replication protein A binding transcriptional activator 1 RBT1.
XX
KW RBT1; replication protein A binding transcriptional activator 1; RPA32;
KW gene therapy; apoptosis; cancer; leukaemia.
XX
OS Unidentified.
XX
PN WO200114546-A2.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-CA000948.
XX
PR 19-AUG-1999; 99US-0149472P.
XX
(TRAN-) CENT TRANSLATIONAL RES IN CANCER.
XX
PI Alaoui-Jamali MA, Cho JM;
XX
DR WPI: 2001-218447/22.
DR N-PSDB; AAF28052.
XX
PT Novel replication protein A binding transcriptional activator 1 gene,
PT useful for treating neoplastic disorders such as cancer and in gene
XX therapy.
XX
PS Disclosure; Fig 1; 16pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC replication protein A binding transcriptional activator 1 (RBT1). The
CC protein is capable of inducing apoptosis. The sequences are useful in the
CC gene therapy and other methods of treatment of cancer, including
CC leukaemias. The present sequence is the RBT1 protein
XX
SQ Sequence 196 AA;

Query Match 100.0%; Score 1033; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.6e-92; Indels 0; Gaps 0;
Matches 196; Conservative 0; Mismatches 0;

QY 1 MEGGLKRKHSLEEEERWESWPAGLQSYQQALLRISLDKVQRSIGPRAPSLRRHVLHN 60
 |||||
 Db 1 MEGGLKRKHSLEEEERWESWPAGLQSYQQALLRISLDKVQRSIGPRAPSLRRHVLHN 60
 |||||
 QY 61 TLQQLAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMGTEPPQNPVTPLG 120
 |||||
 Db 61 TLQQLAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMGTEPPQNPVTPLG 120
 |||||
 QY 121 LQNEVPQDPDPVFLALSSRYLGSGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAPGS 180
 |||||
 Db 121 LQNEVPQDPDPVFLALSSRYLGSGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAPGS 180
 |||||
 QY 181 WENNELDHIMEIILGS 196
 |||||
 Db 181 WENNELDHIMEIILGS 196
 |||||

RESULT 2
 AAM93922
 ID AAM93922 standard; protein; 196 AA.
 XX
 AC AAM93922;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 4085.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94884.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 in genetic manipulation.
 XX
 PS Claim 8; SEQ ID NO 4085; 1380pp + Sequence Listing; English.
 XX

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a polypeptide encoded by a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

Sequence 196 AA;
 Query Match 99.3%; Score 1026; DB 4; Length 196;
 Best Local Similarity 99.5%; Pred. No. 7.9e-92;
 Matches 195; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEGGLKRKHSLEEEERWESWPAGLQSYQQALLRISLDKVQRSIGPRAPSLRRHVLHN 60
 |||||
 Db 1 MEGGLKRKHSLEEEERWESWPAGLQSYQQALLRISLDKVQRSIGPRAPSLRRHVLHN 60
 |||||
 QY 61 TLQQLAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMGTEPPQNPVTPLG 120
 |||||
 Db 61 TLQQLAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMGTEPPQNPVTPLG 120
 |||||
 QY 121 LQNEVPQDPDPVFLALSSRYLGSGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAPGS 180
 |||||
 Db 121 LQNEVPQDPDPVFLALSSRYLGSGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAPGS 180
 |||||
 QY 181 WENNELDHIMEIILGS 196
 |||||
 Db 181 WENNELDHIMEIILGS 196
 |||||

RESULT 3
 ADL32052
 ID ADL32052 standard; protein; 196 AA.
 XX
 AC ADL32052;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE Human protein encoded by a full length cDNA clone SeqID 4085.
 XX
 KW human; medicine; signal transduction; glycoprotein; transcription;
 oligo-capping method.
 XX
 OS Homo sapiens.
 XX
 PN EP1396543-A2.
 XX
 PD 10-MAR-2004.
 XX
 PF 07-JUL-2000; 2003EP-00025638.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183865.
 PR 07-JUL-2000; 2000EP-00114089.
 XX
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2004-204755/20.
 DR N-PSDB; ADL32051.
 XX
 PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
 length human cDNAs.
 XX
 PS Example 1; SEQ ID NO 4085; 1340pp; English.

This invention relates to a novel primers useful for synthesising full length cDNA molecules that encode human proteins. Specifically, it refers to secretory or membrane proteins that are potential therapeutic agents/target molecules in the field of medicine, and in particular genes encoding proteins that are associated with signal transduction, glycoproteins and transcription. The present invention describes a method for efficiently cloning a full length human cDNA from both the 5' and 3' ends using the oligo-capping method. This polypeptide sequence is a full length human protein of the invention.

Sequence 196 AA;
 Query Match 99.3%; Score 1026; DB 8; Length 196;
 Best Local Similarity 99.5%; Pred. No. 7.9e-92;
 Matches 195; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEGGLKRKHSLEEEERWESWPAGLQSYQQALLRISLDKVQRSIGPRAPSLRRHVLHN 60

Db 1 MVGLKRRKHSDEEERWESFAGLSYQQALLRLSLDKVRSIGPRAPSLRRHVLHN 60
QY 61 TLQQQAALRLAPALPEPLFLGEEDFSLSATIGSILRELDTSMDGTEPPQNPVTPLG 120
Db 61 TLQQQAALRLAPALPEPLFLGEEDFSLSATIGSILRELDTSMDGTEPPQNPVTPLG 120
QY 121 LQNEVPQPDVPVLEALSSRYLGDGLDDFFLDIDTSAVEKEPARAPPEPHNLCAPGS 180
Db 121 LQNEVPQPDVPVLEALSSRYLGDGLDDFFLDIDTSAVEKEPARAPPEPHNLCAPGS 180
QY 181 WEWNELDHIMEIILGS 196
Db 181 WEWNELDHIMEIILGS 196

RESULT 4
AAG74742
ID AAG74742 standard; protein; 157 AA.
XX
AC AAG74742;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5506.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
XX Homo sapiens.
OS
XX
XX WO200122920-A2.
FN
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US026524.
PF
XX
XX 29-SEP-1999; 99US-0157137P.
PR
XX 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
XX
DR N-PSDB; AAH34147.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX Claim 11; Page 7098-7099; 9803pp; English.
XX
XX AAH22943 to AAH37195 and AAG73514 to AAG77798 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77799 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
XX Sequence 157 AA;

Query Match 76.6%; Score 791; DB 4; Length 157;
Best Local Similarity 98.0%; Pred. No. 5.4e-69;
Matches 150; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 44 SLGPPAPSLRRHVLHNTLQQLQAALRLAPALPEPLFLGEEDFSLSATIGSILRELD 103
Db 5 ALGPPAPSLRRHVLHNTLQQLQAALRLAPALPEPLFLGEEDFSLSATIGSILRELD 64
QY 104 TSMDCTEPPQNPVTPLGLQNEVPQPDVPVLEALSSRYLGDGLDDFFLDIDTSAVEKEP 163
Db 65 TSMDCTEPPQNPVTPLGLQNEVPQPDVPVLEALSSRYLGDGLDDFFLDIDTSAVEKEP 124
QY 164 ARAPPEPHNLCAPGSWEWNELDHIMEIILGS 196
Db 125 ARAPPEPHNLCAPGSWEWNELDHIMEIILGS 157

RESULT 5
ABO60075
ID ABO60075 standard; protein; 142 AA.
XX
AC ABO60075;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human genome derived single exon protein #6309.
XX
XX Human; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX
XX Homo sapiens.
OS
XX US2003194704-A1.
FN
XX 16-OCT-2003.
PD
XX 03-APR-2002; 2002US-00029386.
PF
XX 03-APR-2002; 2002US-00029386.
PR
XX (PENN/) PENN S G.
PA (RANK/) RANK D R.
PA (HANZ/) HANZEL D K.
XX
XX Penn SG, Rank DR, Hanzel DK;
PI WPI; 2004-119264/12.
DR
XX New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.
XX
XX Claim 45; SEQ ID NO 33709; 80pp; English.
PS
XX The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above, to
CC methods of selling and/or licensing single exon probes or microarrays to

CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterising
 CC alternative splicing events, in detecting and characterising gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe protein of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704
 XX
 SQ Sequence 142 AA;

Query Match 62.5%; Score 645.5; DB 8; Length 142;
 Best Local Similarity 96.8%; Pred. No. 7.8e-55; Indels 1; Gaps 1;
 Matches 122; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 71 LAPAPALPPEPLFLGEEEDFSLSATIGSILRELDTSMDGTPEPPQNPVPLGLQNEVPPQPD 130
 Db 18 LPPLPC-PPEPLFLGEEEDFSLSATIGSILRELDTSMDGTPEPPQNPVPLGLQNEVPPQPD 76
 QY 131 PVFLEALSRYLGDGLDDFFLDIDTSAVEKEPARAPPEPPHNLFCAPGSWEWNELDHIM 190
 Db 77 PVFLEALSRYLGDGLDDFFLDIDTSAVEKEPARAPPEPPHNLFCAPGSWEWNELDHIM 136
 QY 191 EIIIGS 196
 Db 137 EIIIGS 142

RESULT 6
 ABO60074
 ID ABO60074 standard; protein; 111 AA.

AC ABO60074;
 DT 29-JUL-2004 (first entry)
 DE Human genome derived single exon protein #6308.
 XX Human; gene expression; single exon probe; microarray;
 KW alternative splicing event; genomic alteration.
 XX Homo sapiens.

US2003194704-A1.

PD 16-OCT-2003.

PF 03-APR-2002; 2002US-00029386.

PR 03-APR-2002; 2002US-00029386.

PA (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.

PI Penn SG, Rank DR, Hanzel DK;

DR WPI; 2004-119264/12.

PT New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.

PS Claim 45; SEQ ID NO 33708; 80pp; English.
 XX The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridises under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterising
 CC alternative splicing events, in detecting and characterising gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe protein of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704
 XX
 SQ Sequence 111 AA;

Query Match 38.9%; Score 402; DB 8; Length 111;

Best Local Similarity 89.9%; Pred. No. 3.5e-31;

Matches 80; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 MEGGLKRXHSDLEEEERWENSPAGLQSYQALLRISLDKQVRSIGPRAPSLRRHVLHN 60

Db 3 MVGGLKRXHSDLEEEERWENSPAGLQSYQALLRISLDKQVRSIGPRAPSLRRHVLHN 62

QY 61 TLQQLQAALRLAPALPPEPLFLGEEEDF 89

Db 63 TLQQLQAALRLAPALPPEPLPGRGGF 91

RESULT 7

AA36004

ID AAY36004 standard; protein; 236 AA.

XX AAY36004;

DT 13-SEP-1999 (first entry)

XX Extended human secreted protein sequence, SEQ ID NO. 389.

XX Secreted protein; human; cytokine; cellular proliferation; cell movement;
 KW cellular differentiation; immune system regulator; anti-inflammatory;
 KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
 KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
 KW genetic disease.

OS Homo sapiens.

XX WO9931236-A2.

PN

XX

Db 64 LRHLVLVNTLRIQAS--MAPAAALPPVPSPAPASVADNLLASSDAALSASMASLLED 121
QY 102 LDTSMGTEPPQNVPVTLGLQNEVPP-----QDPVFLEAL--SSRYLGDGSLDDPFL 152
Db 122 L-SHIEGLSQAPQ-----LADEGPPGRSIGGAAPSLGALDILGPATGCLLDGLEGUFE 175
QY 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE-----WNELDHIMEIILGS 196
Db 176 DIDTSMYDNLWAPASEGLKPGPED---GPGKEAPELDEALDYLDMDVLVGT 225

RESULT 13
AAM25550
ID AAM25550 standard; protein; 278 AA.
XX
AC AAM25550;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:1065.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiagregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
KW dermatological; anti allergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.

XX Homo sapiens.
XX WO200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US035017.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
XX
PA (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457603/49.
DR N-PSDB; AAH99491.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
PS Claim 20; Page 214; 1217pp; English.
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and cells
CC they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;
CC antiulcer; osteopathic; dermatological; anti allergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis

CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders
XX
SQ Sequence 278 AA;

Query Match 20.6%; Score 213; DB 4; Length 278;
Best Local Similarity 32.6%; Pred. No. 3.4e-12;
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;
QY 4 GLKRXHSLDEEE-----ERWESWAGLOSQQA-----LLRISLDKVSRLGPRAPS 51
Db 47 GLKRXHEEEKEPLAVDSW-WLDFGHAAVAQAPPAVASSSLFDLSVLKHLHSLQQSEPD 105
QY 52 LRHVLHNTLQQLQAALRLAPALPPEPL-----FLGEEDFSLSATIGSILRE 101
Db 106 LRHLVLVNTLRIQAS--MAPAAALPPVPSPAPASVADNLLASSDAALSASMASLLED 163
QY 102 LDTSMGTEPPQNVPVTLGLQNEVPP-----QDPVFLEAL--SSRYLGDGSLDDPFL 152
Db 164 L-SHIEGLSQAPQ-----LADEGPPGRSIGGAAPSLGALDILGPATGCLLDGLEGUFE 217
QY 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE-----WNELDHIMEIILGS 196
Db 218 DIDTSMYDNLWAPASEGLKPGPED---GPGKEAPELDEALDYLDMDVLVGT 267

RESULT 14
AAV31829
ID AAV31829 standard; protein; 236 AA.
XX
AC AAV31829;
XX
DT 06-DEC-1999 (first entry)
XX
DE Human adult blood secreted protein g21_1.
XX
KW Secreted protein; g21_1; human; therapy; diagnosis; vaccine; blood.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /note= "encoded by AWG"
FT Peptide 19..21 /note= "signal peptide"
FT Protein 32..236 /note= "mature protein"
FT Domain 40 /note= "a putative transmembrane domain is centered around this residue"
FT Peptide 67..79 /note= "alternative signal peptide"
FT Protein 80..236 /note= "alternative mature protein"
FT Domain 80 /note= "a putative transmembrane domain is centered around this residue"
FT Misc-difference 137 /note= "encoded by GAS"
FT Domain 150 /note= "a putative transmembrane domain is centered around this residue"
XX WO9947555-A1.
XX 23-SEP-1999.
PD

```
XX PF 18-MAR-1999; 99WO-US005939.
XX XX
XX PR 20-MAR-1998; 98US-0078803F.
XX PR 17-MAR-1999; 99US-00078803.
XX XX
XX PA (GEMY ) GENETICS INST INC.
XX XX
XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Agostino MJ, Steininger RJ;
XX DR WPI; 1999-562059/47.
XX DR N-PSDB; AA219894.
XX XX
XX PT New polynucleotides derived from murine fetal cell cDNA libraries,
XX PT potentially used as, e.g. vaccines.
XX XX
XX PS Claim 13(a); Page 94; 107pp; English.
XX XX
XX CC This is the predicted amino acid sequence of a novel human secreted
XX CC protein, g21_1, as deduced from an isolated adult blood cDNA clone (see
XX CC AA219894). The invention provides new human secreted proteins (see
XX CC AA219894) and polynucleotides (see AA219893-901) isolated from foetal
XX CC cell, adult blood, adult brain and foetal kidney cDNA libraries. They are
XX CC predicted to have biological activities which would make them suitable
XX CC for treating, preventing or ameliorating medical conditions in humans and
XX CC animals, although no supporting data are given. Suggested activities
XX CC include nutritional, cytokine, tissue growth, cell proliferation and
XX CC differentiation, immunostimulant (e.g. as vaccine), immunosuppressive,
XX CC haematopoiesis regulating, activin or inhibin, chemotactic or
XX CC chemokinetic, haemostatic or thrombolytic, receptor/ligand activity,
XX CC antiinflammatory, cadherin or tumour invasion suppressor, and tumour
XX CC inhibition activities
XX CC
XX SQ Sequence 236 AA;
XX
XX Query Match 20.4%; Score 211; DB 2; Length 236;
XX Best Local Similarity 32.6%; Pred. No. 4.3e-12;
XX Matches 76; Conservative 35; Mismatches 70; Indels 52; Gaps 13;
XX
QY 4 GLKRRKHSLEEEE-----ERWESWPAGLSQYQQA-----LLRSLDKVQSLGRAPS 51
DB 5 GLKRRKEEEKPLAVDSW-WLDPGHAAVAQAPPAVASSSLFDLSVLKHHSLQOQSEPD 63
QY 52 LRRHLVHNTLQQAALRLAPALPPEPL-----FLGEEDFSLSATIGSILRE 101
DB 64 LRLHLVNVNLTIRIQAS--MAPAALPPVPPAPPAAPSVADNLLASSDAALSASMASLLED 121
QY 102 LDTSDMGTEPPQNVTPLGLQNEVPP-----QDPVFLEAL--SSRYLGSGLDDPFL 152
DB 122 L-SHTEGLSQAPQ-----LAXEGPPGRSIGGAAPSLGALDILGPATGCLLDGLEGLFE 175
QY 153 DDTSAVEKE---PARAPP-PhnLFCAPGSWE-----WNELDHIMEITLGS 196
DB 176 DDTSDMYNDELWAPASEGLKPGPED---GPGKERAPDELDAELDYMDVLVGT 225
XX
RESULT 15
ABO60367
ID ABO60367 standard; protein; 222 AA.
XX
XX AC ABO60367;
XX
XX DT 29-JUL-2004 (first entry)
XX
XX DE Human genome derived single exon protein #6601.
XX
XX KW Human; gene expression; single exon probe; microarray;
XX KW alternative splicing event; genomic alteration.
XX
XX OS Homo sapiens.
XX
XX PN US2003194704-A1.
```

```
XX PD 16-OCT-2003.
XX
XX PF 03-APR-2002; 2002US-00029386.
XX PR 03-APR-2002; 2002US-00029386.
XX PR
XX PA (PENN/) PENN S G.
XX PA (RANK/) RANK D R.
XX PA (HANZ/) HANZEL D K.
XX
XX PI Penn SG, Rank DR, Hanzel DK;
XX DR WPI; 2004-119264/12.
XX XX
XX PT New human genome-derived single exon nucleic acid probes useful for human
XX PT gene expression analysis, for identifying or characterizing alternative
XX PT splicing events, for assessing genomic alterations or as tools for
XX PT surveying tissues.
XX
XX PS Claim 45; SEQ ID NO 34001; 80pp; English.
XX
XX CC The invention relates to a nucleic acid probe for measuring human gene
XX CC expression, comprising any of the 27,400 fully defined nucleotide
XX CC sequences in the specification, or their complements or fragments, and
XX CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
XX CC fully defined in the specification. The probe is a single exon probe that
XX CC hybridises under high stringency conditions to a nucleic acid molecule
XX CC expressed in human cells or tissues. Also included are a spatially-
XX CC addressable set of single exon nucleic acid probes for measuring human
XX CC gene expression (comprising a plurality of single exon nucleic acid
XX CC probes cited above, where each of the plurality of probes is separately
XX CC and addressably isolatable or amplifiable from the plurality), a single
XX CC exon microarray for measuring human gene expression, a method of
XX CC measuring human gene expression, a vector comprising the single exon
XX CC probe cited above, an ORF-encoded peptide comprising at least 8
XX CC contiguous amino acids of any of the above-mentioned amino acid
XX CC sequences (optionally with conservative amino acid substitutions), an
XX CC isolated antibody that binds specifically to a peptide cited above,
XX CC methods of selling and/or licensing single exon probes or microarrays to
XX CC a customer desiring to measure gene expression, a method of providing
XX CC human gene expression data by subexpression, and a computer-readable
XX CC storage medium which contains a database having a plurality of records
XX CC (each record including data on the expression of a single exon probe
XX CC cited above. The probe, methods and apparatus are useful in gene
XX CC expression analysis. The probes may be used as tools for surveying
XX CC tissues to detect the presence of expressed messages that contain their
XX CC specific exon, or in constructing genome-derived single exon microarrays.
XX CC In addition, the probes are used in identifying and characterising
XX CC alternative splicing events, in detecting and characterising gross
XX CC alterations in the genomic locus that includes their exon, in assessing
XX CC smaller genomic alterations, in priming the synthesis of nucleic acids,
XX CC or in expressing the ORF-encoded peptide. The present sequence is a human
XX CC single exon probe protein of the invention. Note: The sequence data for
XX CC this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from USPTO at
XX CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
XX SQ Sequence 222 AA;
XX
XX Query Match 19.8%; Score 205; DB 8; Length 222;
XX Best Local Similarity 32.6%; Pred. No. 1.5e-11;
XX Matches 75; Conservative 34; Mismatches 69; Indels 52; Gaps 13;
XX
QY 4 GLKRRKHSLEEEE-----ERWESWPAGLSQYQQA-----LLRSLDKVQSLGRAPS 51
DB 5 GLKRRKEEEKPLAVDSW-WLDPGHAAVAQAPPAVASSSLFDLSVLKHHSLQOQSEPD 63
QY 52 LRRHLVHNTLQQAALRLAPALPPEPL-----FLGEEDFSLSATIGSILRE 101
DB 64 LRLHLVNVNLTIRIQAS--MAPAALPPVPPAPPAAPSVADNLLASSDAALSASMASLLED 121
QY 102 LDTSDMGTEPPQNVTPLGLQNEVPP-----QDPVFLEAL--SSRYLGSGLDDPFL 152
```

Db 122 L-SHIEGLSQAPQP-----LADEGPPGRSIGGAPSLGALDILGPA TGCLLDGGLFE 175
Qy 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE-----WNELDHIMEII 193
Db 176 DIDTSMYDNEI WAPASEGLKPGPED---GPGKEAPELDEAELDYIMDVL 222

Search completed: April 25, 2005, 12:29:28
Job time : 170 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2005, 12:23:30 ; Search time 41 Seconds
(without alignments)
459.963 Million cell updates/sec

Title: US-10-069-386A-2
Perfect score: 1033
Sequence: 1 MEGGLKXKHSLEEEERWE.....APGSWEWNELDHIMEIILGS 196

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101.5	9.8	728	2 H59435	phosphoinositide-3
2	97	9.4	803	2 T40514	Chaperonin hsp78p
3	97	9.4	1201	2 G86441	unknown protein [i
4	96	9.3	817	2 S51342	verprolin - yeast
5	95.5	9.2	446	2 A42029	transcription fact
6	93.5	9.1	1734	2 A54602	microtubule-associ
7	93	9.0	433	1 FOLJH2	gag polyprotein -
8	91.5	8.9	1621	2 A82255	hypothetical prote
9	90.5	8.8	804	2 AG0565	probable membrane
10	88	8.5	651	2 T42644	hypothetical prote
11	88	8.5	716	2 T26998	hypothetical prote
12	88	8.5	1676	2 A56508	amuclease primary
13	86.5	8.4	536	2 A34596	transcription fact
14	86.5	8.4	1494	2 T14355	protein-tyrosine-p
15	86	8.3	475	2 B37761	ntrA protein - Thi
16	85.5	8.3	485	2 T37550	hypothetical colle
17	85.5	8.3	825	2 E75508	conserved hypothet
18	85.5	8.3	1544	2 E59431	phosphoinositide-b
19	85	8.2	715	2 D85087	hypothetical prote
20	85	8.2	1217	2 T00270	hypothetical prote
21	84.5	8.2	366	1 SAVLWD	large surface anti
22	84.5	8.2	477	2 T46304	hypothetical prote
23	84.5	8.2	723	2 B38749	3-phosphatidylinos
24	84.5	8.2	908	2 T16057	hypothetical prote
25	84.5	8.2	982	2 T43676	hunchback-related
26	84.5	8.2	1262	2 T13353	protein stn-B - fr
27	84	8.1	289	2 D70452	leucine-tRNA ligas
28	84	8.1	954	2 I61714	co-repressor prote
29	84	8.1	1015	2 JCS062	phogrin precursor

ALIGNMENTS

RESULT 1

H59435
phosphoinositide-3-kinase regulatory beta chain [imported] - human
C:Species: Homo sapiens (man)
C:Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 09-Jul-2004
C:Accession: H59435; A59436
R:Volinia S; Patraccchini P; Otsu M; Hiles I; Gout I; Calzolari E; Bernardi F; Rooke L; W
Oncogene 7, 789-793, 1992
A:Title: Chromosomal localization of human p85 alpha, a subunit of phosphatidylinositol 3-kinase
A:Reference number: H59435
A:Accession: H59435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-728 <VOL>
A:Cross-references: UNIPROT:O00459; GB:NP_005018; PID:G4826908; PIDN:NP_005018.1
R:Janssen, J.W.; Schleithoff, L.; Bartram, C.R.; Schulz, A.S.
Oncogene 16, 1767-1772, 1998
A:Title: An oncogenic fusion product of the phosphatidylinositol 3-kinase p85beta subunit
A:Reference number: A59436; MUID:98241181; PMID:9582025
A:Accession: A59436
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-728 <JAN>
A:Cross-references: GB:NP_005018; PID:G4826908; PIDN:NP_005018.1

Query Match 9.8%; Score 101.5; DB 2; Length 728;
Best Local Similarity 23.9%; Pred No. 1;
Matches 47; Conservative 25; Mismatches 70; Indels 55; Gaps 9;

QY	20	EWSPA----	GLQSYQQAL-----	LRISLDKVQVSLGPRAPSLRRHLIH	59
DB	164	QWDTAALADGKISFLIALPALPALTPEASARRALREAAAGVGPALPEPTLPLHRAALTJR	223		
QY	60	NTLQQLAALRLAPALPFLPELFLGEEDFSLSATIGLSILRELDFTSMGTGPQNPVPL	119		
DB	224	FLQLHLGRVARRAPA-----	LGPAVALGATFGFL-----	LRAPPPSPSP-PFG	267
QY	120	GLQNEVPQPPD--PVFLEALSSRYLGDSGLDDFDIDTSAVEKEPARAPPP-----	170		
DB	268	GAPGSESPDPFALLVEKLQEHLEQ-----	EVAPPALPPKPKAKPAFTVLANGG	320	
QY	171	-PHNLFCAPGSWEWNEEL	186		
DB	321	SPPSL--QDAEWYMGDI	335		

RESULT 2

T40514
Chaperonin hsp78p - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T40514

Qy 96 GSILRELDTSMDGTEPPQNPV-----TPGLQNEVP---POPDPVFLEALSRYL 142

```
Db 634 ASLLHDVLFSSNLEKPPSPVAAAPLPTFSAPSLQQSVSTSIQPPPV-APTLSVR-- 690
Qy 143 GDSGLDDFFLDIDTSAVEKEPARAPPP 171
Db 691 -----TETESISKNTKSRPPPP 708

RESULT 5
A42029
transcription factor E3 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A42029
R:Roman, C.; Matera, A.G.; Cooper, C.; Artandi, S.; Blain, S.; Ward, D.C.; Calame, K.
Mol. Cell. Biol. 12, 817-827, 1992
A:Title: mTPE3, an X-linked transcriptional activator containing basic helix-loop-helix
A:Reference number: A42029; MUID:92123207; PMID:1732746
A:Accession: A42029
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-446 <ROM>
A:Cross-references: UNIPROT:064092; GB:S76673; PIDN:AAB21130.1; PID:G243440
A:Note: sequence extracted from NCBI backbone (NCBIN:76673, NCBI:76674)

Query Match 9.2%; Score 95.5; DB 2; Length 446;
Best Local Similarity 27.3%; Pred. No. 1.7;
Matches 48; Conservative 24; Mismatches 61; Indels 43; Gaps 8;

Qy 12 LEEEREWSPAGLQSQVQALLRLSLDKVQSLGPRAPSLRRHVLHNTLQQLAALRL 71
Db 272 LQKQQR-----SKQLRSR-----SLEQANSLQLRIQEL-----ELQRIHG 311
Qy 72 APAPALPPPLFLGDEDFSLATIGSILRELDTSMDGTPEPPQNPVTPLGLQNEVPPQDP 131
Db 312 LP---VPPNPGLSLTSSVSDSLKP--RQDIIEEGRPSTTFHVSQGPAAQNPQQPPA 366
Qy 132 VFLEAL-----SSRYLGDGLDDFFLDIDTSAVEKE-----PARAPPEP 170
Db 367 PPSDALLDLHPFSDHLGDLG-DPFLHGLRDLIMEBEGVGLSGGALSPLRAASDP 421

RESULT 6
A54602
microtubule-associated serine/threonine protein kinase MAST205 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: A54602
R:Walden, P.D.; Cowan, N.J.
Mol. Cell. Biol. 13, 7625-7635, 1993
A:Title: A novel 205-kilodalton testis-specific protein kinases; protein kinase homolog
A:Reference number: A54602; MUID:94067123; PMID:8246979
A:Accession: A54602
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1734 <WAL>
A:Cross-references: UNIPROT:060592; GB:U02313; NID:G406057; PIDN:AAC04312.1; PID:G406058
A:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
A:Keywords: ATP
F:451-726/Domain: protein kinase homology <KIN>
F:459-467/Region: protein kinase ATP-binding motif

Query Match 9.1%; Score 93.5; DB 2; Length 1734;
Best Local Similarity 26.3%; Pred. No. 15;
Matches 51; Conservative 23; Mismatches 71; Indels 49; Gaps 8;

Qy 25 GLQSQVQAL-----LRISLDKVQSLGPRAPSLRRHVLHNTLQQLAALRLAPALPP 79
Db 1308 GSQSPFTKLHSPFLGRQLSRPKSAEPPRSPLLKR-----VQSAEKLAALAAAE 1357
Qy 80 EPLFLGEEDFSLSATIGSLRELDTSMDGTPEPPQNPVTPLGLQNEVP---PPQDPVFLEA 136
Db 1358 KKL-APSRKHSLLDPHGELKKEL-----TPREASPLEVVGTRSVLSGKGLPGKVLQP 1410
```

```
Qy 137 LSSRYLG-----DSGLDDFFLDIDTSAVEKEPARAP-PPPPHN 173
Db 1411 AFSRALGTLRQDRAERRESLQKEAIRVEDSDDTDEEPNSQATQEPRLSPHPPEASHN 1470
Qy 174 LFCAPGSWEWNELD 187
Db 1471 LL-PKSGEGTEED 1493
```

RESULT 7
FOLJH2

gag polyprotein - human T-cell lymphotropic virus type 2

N:Alternate names: core polyprotein

N:Contains: core protein p12; core protein p15; core protein p24

C:Species: human T-cell lymphotropic virus type 2, HTLV-2

A:Note: host Homo sapiens (man)

C:Date: 28-Feb-1986 #sequence_revision 28-Feb-1986 #text_change 09-Jul-2004

C:Accession: A03944

R:Shimotohno, K.; Takahashi, Y.; Shimizu, N.; Gojobori, T.; Golde, D.W.; Chen, I.S.Y.; Mi

Proc. Natl. Acad. Sci. U.S.A. 82, 3101-3105, 1985

A:Title: Complete nucleotide sequence of an infectious clone of human T-cell leukemia vi

A:Reference number: A94042; MUID:85216449; PMID:2582407

A:Accession: A03944

A:Molecule type: DNA

A:Residues: 1-433 <SHI>

A:Cross-references: UNIPROT:P03346; GB:M10060; NID:G329559; PIDN:AAB59884.1; PID:G329562

A:Note: the authors translated the codon TAC for residues 197 and 249 as Thr

C:Genetics:

A:Gene: gag

C:Superfamily: mammalian retrovirus gag polyprotein II

C:Keywords: core protein; polyprotein

F:1-136/Product: core protein p15 #status predicted <P15>

F:137-214/Product: core protein p24 #status predicted <P24>

F:215-433/Product: core protein p12 #status predicted <P12>

Query Match 9.0%; Score 93; DB 1; Length 433;

Best Local Similarity 23.1%; Pred. No. 2.7;

Matches 40; Conservative 24; Mismatches 67; Indels 42; Gaps 7;

Qy 47 PRAPSLRRHVLHNTLQQLAALRLAPAPA-----LPPEPLFGEEDFSLA 93

Db 13 PRAP--RGLSTHWHVLFQAAVRLQPRPSDFDQQLRRFLKALKTFWLAPIDYSLA 69

Qy 94 TT-----GSILRELDTSMDGTPEPPQNPVTPLG---LQNEVPPQDPVFLEA-LSSRYLG 143

Db 70 SLIPKGYGRVVEIINILKQNVSPAPAPVPTPICPTTTTPPPPPSPPAHVPPPY-- 127

Qy 144 DSGLDDFFLDIDTSAVEKEPARAPPEPPHNLFCAPGSWEWNELDHIMEILGS 196

Db 128 -----VEPTTTCQFPILHPPGAP-----SAHRPMQMKDLOAIKQEVSS 166

RESULT 8

A82255

hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

R:Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: A82255

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1621 <HRI>

A:Cross-references: UNIPROT:Q9KTA5; GB:AE004181; GB:AE003852; NID:G9655454; PIDN:AAF9415

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0998

R;Cao, L.; Zhang, L.; Ruiz-Lozano, P.; Yang, Q.; Chien, K.R.; Graham, R.M.; Zhou, M.
J. Biol. Chem. 273, 21077-21083, 1998
A;Title: A novel putative protein-tyrosine phosphatase contains a BRO1-like domain and st
A;Reference number: Z18004; MUID:98361981; PMID:9694860
A;Accession: T14355
A;Status: preliminary; translated from GB/EMBL/DBDJB
A;Molecule type: mRNA
A;Residues: 1-1494 <CAO>
A;Cross-references: UNIPROT:O88902; EMBL:AF077000; NID:g3598973; PID:g3598974; PIDN:AAC6:
A;Experimental source: brain
C;Genetics:
A;Gene: ptp-TD14
C;Function:
A;Description: may be involved in regulating Ha-ras-dependent cell growth
C;Keywords: phosphoric monoester hydrolase

Query Match 8.4%; Score 86.5; DB 2; Length 1494;
Best Local Similarity 29.1%; Pred. No. 48;
Matches 37; Conservative 19; Mismatches 40; Indels 31; Gaps 9;

QY 62 LQQLQAALRLAPALPPEFLFGEEDFSLSATIGSILRE-----LDTSMDGTEPPNPQ 115
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||
Db 625 LMQPRAAVMAPGVLVPAPVVTSE-----LGLVRSSPQGIVSSPYAGVGPPQ-P 675

QY 116 VTPLGLQNEVPPQ---PDVFLEALSRYLGDSGLDDFFLDIDT-SAVEKEPARAPPPEP 171
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||
Db 676 I--VGLFSAPPQFGSGE----LAMDVLR-PATTTVDVSQVAPISHHMALRPGAPAPPQP- 727

QY 172 HNLFCAP 178
Db 728 ----CFP 730

RESULT 15
B37761
ntrA protein - Thiobacillus ferrooxidans
C;Species: Thiobacillus ferrooxidans
C;Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 09-Jul-2004
C;Accession: B37761
R;Berger, D.K.; Woods, D.R.; Rawlings, D.E.
J. Bacteriol. 172, 4399-4406, 1990
A;Title: Complementation of Escherichia coli sigma(54) (ntrA)-dependent formate hydrogen
A;Reference number: A37761; MUID:90330545; PMID:2198257
A;Accession: B37761
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-475 <BER>
A;Cross-references: UNIPROT:P24695; GB:M58480; GB:M33831; NID:g154642; PIDN:AAA27379.1; I:
C;Superfamily: Pseudomonas transcription initiation factor sigma
C;Keywords: DNA binding; transcription regulation

Query Match 8.3%; Score 86; DB 2; Length 475;
Best Local Similarity 21.7%; Pred. No. 12;
Matches 40; Conservative 31; Mismatches 75; Indels 38; Gaps 6;

QY 1 MEGGLRKHSDLDEERWESWPSAGLSQYQALLRI SLDKVQR-----SLGPRA--PSLR 53
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||
Db 307 MAGCKDAAHKYIQDLNEARWFIKSLGRQDTILKVARAI VERQKDFPFANGPESMRPMVL 366

QY 54 RHVLIHNTLQOLQAURLAPALPPEFLFGEEDFSLSATIG-----STIRELDT 104
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||
Db 367 RH--IADAVEMHESTSVRTVNQKYMTPRGLYEPKYPFSSHVGTSDGGSASATAIRALI 424

QY 105 SMDGTBPQNPTPLGLQNEVPPQDPVFLEALSRYLGDSGLDDFFLDIDTSAVEKEPA 164
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||
Db 425 KMTQAEADAQHPLS-----DAETARVLADQGIQ---IARTTVAKYREA 464

QY 165 RAPP 168
Db 465 NVPP 468

Search completed: April 25, 2005, 12:33:15

Search completed: April 25, 2005, 12:33:15

Job time : 44 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2005, 12:32:36 ; Search time 134 Seconds
(without alignments)
486.769 Million cell updates/sec

Title: US-10-069-386A-2

Perfect score: 1033

Sequence: 1 MEGSLKRXHSLREERWE.....APGSWEWELDHIMEILLGS 196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1424015 seqs, 332791073 residues

Total number of hits satisfying chosen parameters: 1424015

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgm2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgm2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
4: /cgm2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgm2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgm2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgm2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgm2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgm2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgm2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgm2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgm2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgm2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgm2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgm2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgm2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgm2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgm2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgm2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgm2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	791	76.6	157	14	US-10-106-698-5516
2	645.5	62.5	142	14	Sequence 5516, Ap
3	402	38.9	111	14	Sequence 33709, A
4	216	20.9	236	11	US-10-029-386-33708
5	213	20.6	236	9	Sequence 33708, A
6	213	20.6	236	11	US-09-978-360A-568
7	213	20.6	236	10	Sequence 568, App
8	205	19.8	222	15	US-09-731-872-412
9	136	13.2	237	13	Sequence 412, App
10	125	12.1	314	15	US-09-876-997-412
11	125	12.1	314	16	Sequence 1065, Ap
12	106	10.3	254	9	US-10-029-386-34001
13	106	10.3	254	10	Sequence 34001, A
					Sequence 2, Appli
					Sequence 31, Appl
					Sequence 627, App
					Sequence 596, App
					Sequence 536, App

14	106	10.3	254	14	US-10-106-698-4456	Sequence 4456, Ap
15	105	10.2	241	13	US-10-076-069-4	Sequence 4, Appli
16	104.5	10.1	740	15	US-10-276-774-2083	Sequence 2083, Ap
17	98	9.5	562	15	US-10-104-047-3214	Sequence 3214, Ap
18	97	9.4	803	15	US-10-369-493-2536	Sequence 2536, Ap
19	96	9.3	297	15	US-10-425-114-64002	Sequence 64002, A
20	96	9.3	817	15	US-10-369-493-1813	Sequence 1813, Ap
21	95.5	9.2	1228	16	US-10-437-963-188300	Sequence 188300, Ap
22	95	9.2	713	16	US-10-437-963-137248	Sequence 137248, Ap
23	94	9.1	433	8	US-08-259-451-3	Sequence 3, Appli
24	94	9.1	433	14	US-10-224-999A-3472	Sequence 3472, Ap
25	93.5	9.1	256	15	US-10-421-138A-312	Sequence 312, App
26	93.5	9.1	256	15	US-10-374-780A-1247	Sequence 1247, Ap
27	93.5	9.1	663	16	US-10-755-889-480	Sequence 480, App
28	93.5	9.1	1734	9	US-09-862-027-81	Sequence 81, Appli
29	93.5	9.1	1734	15	US-10-042-865-82	Sequence 82, Appli
30	93	9.0	157	16	US-10-767-701-54984	Sequence 54984, A
31	92.5	9.0	113	13	US-10-076-069-6	Sequence 6, Appli
32	92	8.9	1103	15	US-10-437-963-166325	Sequence 166325, A
33	91.5	8.9	714	15	US-10-425-114-62753	Sequence 62753, A
34	91	8.8	316	15	US-10-425-114-56782	Sequence 56782, A
35	90.5	8.8	206	16	US-10-437-963-187376	Sequence 187376, A
36	90.5	8.8	370	14	US-10-192-381-16	Sequence 16, Appli
37	90.5	8.8	824	14	US-10-226-844-1	Sequence 1, Appli
38	90.5	8.8	824	14	US-10-210-951-58	Sequence 58, Appli
39	90.5	8.8	824	14	US-10-211-884-58	Sequence 58, Appli
40	90.5	8.8	824	15	US-10-211-858-58	Sequence 58, Appli
41	90.5	8.8	824	16	US-10-789-378-4	Sequence 4, Appli
42	90.5	8.8	824	17	US-10-741-600-1079	Sequence 1079, Ap
43	90.5	8.8	947	9	US-09-871-889-1	Sequence 1, Appli
44	90.5	8.8	947	10	US-09-981-397A-18	Sequence 18, Appli
45	90.5	8.8	947	13	US-10-087-192-888	Sequence 888, App

ALIGNMENTS

RESULT 1

US-10-106-698-5516
; Sequence 5516, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5516
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (55)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (132)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5516

Query Match 76.6%; Score 791; DB 14; Length 157;

Best Local Similarity 98.0%; Pred. No. 9,2e-64;

Matches 150; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY

44 SLGPRAPSLRRHVLHNTLQQLQAALRLAPALPEPLFGEEDFSLSATIGSLRELD 103

Db 5 ALGPAPSLRRHVLHNTLQQAALRLAPALPPEPLFLGEEDFSLXAGISILRELD 64
QY 104 TSMGTEPPQNPVTPLGLQNEVPPQDPVFLBALSRYLGSGLDDFFLDIDTSAVEKEP 163
Db 65 TSMGTEPPQNPVTPLGLQNEVPPQDPVFLBALSRYLGSGLDDFFLDIDTSAVEKEP 124
QY 164 ARAPPEPHNLFPCAGSWEWNELDHIMELIIGS 196
Db 125 ARAPPEPHNLFPCAGSWEWNELDHIMELIIGS 157
RESULT 2
US-10-029-386-33709
; Sequence 33709, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33709
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010271.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.64
US-10-029-386-33709
Query Match 62.5%; Score 645.5; DB 14; Length 142;
Best Local Similarity 96.8%; Pred. No. 1.3e-50;
Matches 122; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 71 LAPALPPEPLFLGEEDFSLXAGISILRELDTSMGTEPPQNPVTPLGLQNEVPPQDP 130
Db 18 LPPLPC-PPEPLFLGEEDFSLXAGISILRELDTSMGTEPPQNPVTPLGLQNEVPPQDP 76
QY 131 PVFLBALSRYLGSGLDDFFLDIDTSAVEKEPARAPPEPHNLFPCAGSWEWNELDHIM 190
Db 77 PVFLBALSRYLGSGLDDFFLDIDTSAVEKEPARAPPEPHNLFPCAGSWEWNELDHIM 136
QY 191 EIILGS 196
Db 137 EIILGS 142
RESULT 3
US-10-029-386-33708
; Sequence 33708, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33708
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

; OTHER INFORMATION: MAP TO AC010271.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.64
US-10-029-386-33708
Query Match 38.9%; Score 402; DB 14; Length 111;
Best Local Similarity 89.9%; Pred. No. 1.1e-28;
Matches 80; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 MEGGLKRXHSDLEEEERWESWSPAGLOSQQALLRISLDKVQRSIGPPAPSLRRHVLHNN 60
Db 3 MVGGLKRXHSDLEEEERWESWSPAGLOSQQALLRISLDKVQRSIGPPAPSLRRHVLHNN 62
QY 61 TLQQLQAALRLAPALPPEPLFLGEEDF 89
Db 63 TLQQLQAALRLAPALPPEPLFLGEEDF 91
RESULT 4
US-09-978-360A-568
; Sequence 568, Application US/09978360A
; Publication No. US20040110939A1
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; APPLICANT: Clusel, Catherine
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides
; FILE REFERENCE: 56.USA.CIP
; CURRENT APPLICATION NUMBER: US/09/978.360A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/066,677
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: US 60/069,957
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: US 60/074,121
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: US 60/081,563
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: US 60/096,116
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: US 60/099,273
; PRIOR FILING DATE: -09-04
; PRIOR APPLICATION NUMBER: US 09/191,997
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: US 09/215,435
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: PCT/IB98/02122
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: US 09/247,155
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 810
; SOFTWARE: Patent.pm
; SEQ ID NO 568
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -31...-1
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28,30,40,67,86,117,120)
; OTHER INFORMATION: unknown
US-09-978-360A-568
Query Match 20.9%; Score 216; DB 11; Length 236;
Best Local Similarity 32.6%; Pred. No. 2.1e-11;
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;
QY 4 GLKRXHSDLEEE-----ERWESWSPAGLOSQQO-----LLRISLDKVQRSIGPPAPSLRRHVLHNN 51

Db 5 GLKRRKEEKEPLAVDSW-WLDPGHAQAQPAVASSSLFDLSVLKHHSLQXSPD 63
QY 52 LRRHVLHNTLQOQAALRLAPALPPEPL-----FLGEEDFSLSATIGSIILRE 101
Db 64 LRHLVLVNTLRRIQAS--MAPAALPPVPPTPPAAPKVADNLLASSDAALSASMAXILLED 121
QY 102 LDTSDMGTEPPQNPVTPLGLQNEVPPQPD-----PVPFLEAL-----SSRYLGDGSLDDPFL 152
Db 122 L-SHIEGLSQAPQ-----LADEGPPGRSIGXPPXGALDGLGATGCLLDNGLEGLFE 175
QY 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE-----WNELDHIMEIILGS 196
Db 176 DIDTSMYDNLWAPASEGLKPGPED---GPGKEAPELDEALDYLDMDVLVGT 225

RESULT 5

US-09-731-872-412
; Sequence 412, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 412
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-731-872-412

Query Match 20.6%; Score 213; DB 9; Length 236;
Best Local Similarity 32.6%; Pred. No. 3.9e-11;
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;
QY 4 GLKRRKHSLEEE-----ERWESPAQLOSYQQA-----LLRISLDKVORSIGPRAPS 51
Db 5 GLKRRKEEKEPLAVDSW-WLDPGHTAVAQAQPAVASSSLFDLSVLKHHSLQXSEPD 63
QY 52 LRRHVLHNTLQOQAALRLAPALPPEPL-----FLGEEDFSLSATIGSIILRE 101
Db 64 LRHLVLVNTLRRIQAS--MAPAALPPVPSPAAPSVADNLLASSDAALSASMAXILLED 121
QY 102 LDTSDMGTEPPQNPVTPLGLQNEVPP-----QPDVPFLEAL-----SSRYLGDGSLDDPFL 152
Db 122 L-SHIEGLSQAPQ-----LADEGPPGRSIGAAPSLGALDGLGATGCLLDNGLEGLFE 175
QY 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE-----WNELDHIMEIILGS 196
Db 176 DIDTSMYDNLWAPASEGLKPGPED---GPGKEAPELDEALDYLDMDVLVGT 225

RESULT 6

US-09-876-997-412
; Sequence 412, Application US/09876997
; Publication No. US20030152921A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US4.CIP
; CURRENT APPLICATION NUMBER: US/09/876,997
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 09/731,872

; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 412
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-876-997-412

Query Match 20.6%; Score 213; DB 10; Length 236;
Best Local Similarity 32.6%; Pred. No. 3.9e-11;
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;
QY 4 GLKRRKHSLEEE-----ERWESPAQLOSYQQA-----LLRISLDKVORSIGPRAPS 51
Db 5 GLKRRKEEKEPLAVDSW-WLDPGHTAVAQAQPAVASSSLFDLSVLKHHSLQXSEPD 63
QY 52 LRRHVLHNTLQOQAALRLAPALPPEPL-----FLGEEDFSLSATIGSIILRE 101
Db 64 LRHLVLVNTLRRIQAS--MAPAALPPVPSPAAPSVADNLLASSDAALSASMAXILLED 121
QY 102 LDTSDMGTEPPQNPVTPLGLQNEVPP-----QPDVPFLEAL-----SSRYLGDGSLDDPFL 152
Db 122 L-SHIEGLSQAPQ-----LADEGPPGRSIGAAPSLGALDGLGATGCLLDNGLEGLFE 175
QY 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE-----WNELDHIMEIILGS 196
Db 176 DIDTSMYDNLWAPASEGLKPGPED---GPGKEAPELDEALDYLDMDVLVGT 225

RESULT 7

US-10-296-115-1065
; Sequence 1065, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1065
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-1065

Query Match 20.6%; Score 213; DB 15; Length 278;
Best Local Similarity 32.6%; Pred. No. 4.8e-11;
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;
QY 4 GLKRRKHSLEEE-----ERWESPAQLOSYQQA-----LLRISLDKVORSIGPRAPS 51
Db 47 GLKRRKEEKEPLAVDSW-WLDPGHAQAQPAVASSSLFDLSVLKHHSLQXSEPD 105
QY 52 LRRHVLHNTLQOQAALRLAPALPPEPL-----FLGEEDFSLSATIGSIILRE 101
Db 106 LRHLVLVNTLRRIQAS--MAPAALPPVPSPAAPSVADNLLASSDAALSASMAXILLED 163
QY 102 LDTSDMGTEPPQNPVTPLGLQNEVPP-----QPDVPFLEAL-----SSRYLGDGSLDDPFL 152
Db 164 L-SHIEGLSQAPQ-----LADEGPPGRSIGAAPSLGALDGLGATGCLLDNGLEGLFE 217
QY 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE-----WNELDHIMEIILGS 196


```
QY 146 GLDD-PFLDIDTSAVEKEP-----ARAPPEPHNLF-----CAPGSWEW 183
Db 242 TLUDDLFAIDITSMYDFDPTCTSSGTASKAPVSAADLLKTLAPYSSQVPTPSQPFKMDL 301
QY 184 NELDHIMEIILGS 196
Db 302 TELDHIMEVLVGS 314

RESULT 11
US-10-755-889-627
; Sequence 627, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 627
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-627

Query Match 12.1%; Score 125; DB 16; Length 314;
Best Local Similarity 22.7%; Pred. No. 0.0053;
Matches 71; Conservative 34; Mismatches 88; Indels 120; Gaps 12;

QY 2 EGLKRGKSHDLSEEREWSPAGLQ-----YQALLRISLDKQVRSGLGPRAPSLRRH 55
Db 4 KGG-KRKDFEHDGLEGKIVSPCDGPSKVSYTQRTQIFNISLMKLYNHRPLTEPSLQKT 62
QY 56 VLIHNTLOQLQAALR-----PALPPE---PLFLGEEDFSLSATIGSLRELD 103
Db 63 VLIINWLRRIQELKQESLRPMFTPSQPTTEPSDSYREAPPASHLASPSHPCDILGS 122
QY 71 -----LAPA-----PALPPE---PLFLGEEDFSLSATIGSLRELD 103
Db 123 TTPLEACLTPALSLEDDDDTFTCSQAMQPTATKLSPPALLPEKD-SFSSALDEIEELCP 181
QY 104 TSMD-----GTEPPQNVTPLGLQNEVPPQ-----PDVFLEAL-----SSRYLGDS 145
Db 182 TSTSTEAATAATSDSVKGTSSSEAGTKLDGQPSRADDSKLMDSLFCNFETITTTGFLTDL 241
QY 146 GLDD-PFLDIDTSAVEKEP-----ARAPPEPHNLF-----CAPGSWEW 183
Db 242 TLUDDLFAIDITSMYDFDPTCTSSGTASKAPVSAADLLKTLAPYSSQVPTPSQPFKMDL 301
QY 184 NELDHIMEIILGS 196
Db 302 TELDHIMEVLVGS 314

RESULT 12
US-09-925-302-596
; Sequence 596, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-596

Query Match 10.3%; Score 106; DB 10; Length 254;
Best Local Similarity 25.6%; Pred. No. 0.21;
Matches 65; Conservative 32; Mismatches 75; Indels 82; Gaps 13;

QY 4 GLKRGKSHDLSEEREWSPAGLQ---SY---QQALLRISLDKQVRSGLGPRAPSLRRHVL 57
Db 18 GLKRGKCVGHEEDV---EGALAGLKTIVSSYSLQSQSLDMSLVKLQCHMLVFNLCRSVL 74
QY 58 IHNTLQQLQAAL-----RLAP---APALPPPEPLFLGEBDFSLSATIGSLRELDTSMDG 108
Db 75 IANTVRQIQEEMTQDGTWRTVAPQAARAPXDRLVSTE-----ILCEAAWGQSG 123
QY 109 TEP-----PQNPT-----PLGLQNEVPPQDP-----VFLEALS 138
Db 124 AHPAPGLGDTGTPGVSDLCPTVSAQAPRHQLSSAWEMDGPENRGSPFKSLDQIPETLE 183
QY 139 SRYLGDSGLDDFFLDIDTSAVEKEP-----AR-----APPEPHNLFCAQGS 180
Db 184 TK--NPSCMEELFSDVDSFYDLDLTVLGTMMGARGPCEGLEGLAPATPGPSSCKS-- 239
QY 181 WENNELDHIMEIIL 194
Db 240 -DLGELDHVVELV 252

RESULT 13
US-09-925-302-596
; Sequence 596, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 596
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (105)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-596

Query Match 10.3%; Score 106; DB 10; Length 254;
Best Local Similarity 25.6%; Pred. No. 0.21;
Matches 65; Conservative 32; Mismatches 75; Indels 82; Gaps 13;

QY 4 GLKRGKSHDLSEEREWSPAGLQ---SY---QQALLRISLDKQVRSGLGPRAPSLRRHVL 57
Db 18 GLKRGKCVGHEEDV---EGALAGLKTIVSSYSLQSQSLDMSLVKLQCHMLVFNLCRSVL 74
```


GenCore version 5.1.1.6.
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2005, 12:24:20 ; Search time 42 Seconds
(without alignments)
348.362 Million cell updates/sec

Title: US-10-069-386A-2

Perfect score: 1033

Sequence: 1 MEGGLKRXHSLREERWE.....APGSWEWELDHIMEIILGS 196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCITUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95.5	9.2	529	4	US-09-949-016-8918
2	94	9.1	433	4	US-08-259-451-3
3	93	9.0	433	1	US-07-672-483-4
4	93	9.0	2618	3	US-09-413-814-28
5	92.5	9.0	432	5	PCT-US95-04910-13
6	92	8.9	379	4	US-09-949-016-10257
7	90.5	8.8	370	4	US-09-377-285B-16
8	90.5	8.8	947	2	US-08-887-518-2
9	90.5	8.8	947	2	US-09-023-321-2
10	90.5	8.8	947	2	US-09-032-475-2
11	90.5	8.8	947	3	US-09-257-703-1
12	90.5	8.8	947	4	US-09-871-889A-1
13	90.5	8.8	1027	4	US-09-902-540-11750
14	89.5	8.7	215	4	US-08-778-717-9
15	89.5	8.7	611	4	US-09-949-016-7139
16	88.5	8.6	897	4	US-09-849-602-18
17	88	8.5	1130	4	US-09-976-594-280
18	87.5	8.5	709	4	US-10-118-328-4
19	86.5	8.4	136	4	US-08-259-451-5
20	85	8.2	1148	4	US-09-949-016-6798
21	84	8.1	337	4	US-09-543-681A-7444
22	84	8.1	514	4	US-09-252-991A-25281
23	84	8.1	628	4	US-09-345-473E-48
24	84	8.1	675	4	US-09-949-016-11490
25	84	8.1	1015	4	US-09-949-016-6276
26	84	8.1	2152	3	US-09-036-987A-3
27	84	8.1	2152	3	US-09-370-700-3

28	84	8.1	2152	4	US-09-603-207-3	Sequence 3, Appli
29	83	8.0	969	2	US-08-548-159-1	Sequence 1, Appli
30	83	8.0	986	2	US-08-548-159-3	Sequence 3, Appli
31	83	8.0	1012	3	US-08-811-481-16	Sequence 16, Appl
32	83	8.0	1012	4	US-09-876-527-16	Sequence 16, Appl
33	82.5	8.0	1130	4	US-09-538-092-834	Sequence 834, App
34	82.5	8.0	1169	4	US-09-949-016-9630	Sequence 9630, Ap
35	82.5	8.0	2142	4	US-09-538-092-1142	Sequence 1142, Ap
36	82	7.9	667	4	US-09-248-796A-18663	Sequence 18663, A
37	82	7.9	1006	4	US-09-023-905A-12	Sequence 12, Appl
38	82	7.9	2101	1	US-08-466-390-4	Sequence 4, Appli
39	82	7.9	2101	1	US-08-470-950-4	Sequence 4, Appli
40	82	7.9	2101	1	US-08-467-781-4	Sequence 4, Appli
41	82	7.9	2101	1	US-08-195-487-4	Sequence 4, Appli
42	82	7.9	2101	2	US-08-483-924-4	Sequence 1, Appli
43	82	7.9	2101	3	US-09-452-294-1	Sequence 4, Appli
44	82	7.9	2101	5	PCT-US93-06160-4	Sequence 4, Appli
45	82	7.9	2107	4	US-09-949-016-7646	Sequence 7646, Ap

ALIGNMENTS

RESULT 1

US-09-949-016-8918

; Sequence 8918, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CLO01307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8918

; LENGTH: 529

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-8918

Query Match

Best Local Similarity 9.2%; Score 95.5; DB 4; Length 529;

Matches 46; Conservative 25; Mismatches 71; Indels 55; Gaps 9;

QY 20 EWSPA---GLQYQOAL-----LRISLDKVQSLGPRAPSLRRHLVLIH 59

Db 174 QWDTAALADGKISFLIALPAPLVTPESAAARRAUREAGVGPALPEPTLPHRALTLR 233

QY 60 NTLQQLQAALRLAPALPPEPLFGEEDFSLSATIGSILRELDTSMDGTEPPQNPVTP 119

Db 234 FLQLQHLGRVASRAPE-----LGPVAVLGAITFGPLL-----LRAPPPSSP-P 277

QY 120 GLQNEVPQPD--PVFLEALSRVYLGDSGLDDFDIDTSVAKEKPARAPPP----- 170

Db 278 GAPDGESESPDFFALLVKEKLQEHLEEQ-----EVAPPALPPKPKAKPAPTVLANG 330

QY 171 -PHNLFCAPGSWEWNEEL 186

Db 331 SPSSL--QDAEWYMGDI 345

RESULT 2

US-08-259-451-3

; Sequence 3, Application US/08259451

; Patent No. 6406841

; GENERAL INFORMATION:
 ; APPLICANT: Lee, Helen H.
 ; APPLICANT: Swanson, Priscilla A.
 ; APPLICANT: Idler, Kenneth B.
 ; APPLICANT: Rosenblatt, Joseph D.
 ; APPLICANT: Chen, Irvin S. Y.
 ; APPLICANT: Golde, David W.
 ; APPLICANT: Robertson, Eugene F.
 ; APPLICANT: Stephens, John E.
 ; APPLICANT: Chan, Emerson W.
 ; APPLICANT: Buytendorp, Mark H.
 ; APPLICANT: Johnson, Joan E.
 ; APPLICANT: Motley, Cheryl T.
 ; APPLICANT: Peterson, Bryan
 ; APPLICANT: Edwards, Michelle
 ; APPLICANT: Guidinger, Peggy
 ; APPLICANT: Tate, Cynthia
 ; TITLE OF INVENTION: HTLV-1/INRA Compositions
 ; TITLE OF INVENTION: and Assays for Detecting HTLV Infection
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Abbott Laboratories
 ; STREET: One Abbott Park Road
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WordPerfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/259,451
 ; FILING DATE: 20-JUN-1994
 ; CLASSIFICATION: 436
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/086,415
 ; FILING DATE: 01-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Daniel W. Collins
 ; REGISTRATION NUMBER: 31,912
 ; REFERENCE/DOCKET NUMBER: 5381.US.P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (708) 937-6365
 ; TELEFAX: (708) 938-2623
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 433 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: unknown
 ; US-08-259-451-3

Query Match 9.1%; Score 94; DB 4; Length 433;
 Best Local Similarity 23.1%; Pred. No. 0.71;
 Matches 40; Conservative 24; Mismatches 67; Indels 42; Gaps 7;
 QY 47 PRAPSLRRHVLHNTLQQLAALRLAPAPA-----LPPEPLFLGEEDFSLSA 93
 DB 13 PKAP---RGLSTHHLNLFQAAYRLQRPSPDFQQLRRFLKALKTPILNLPIDYSLA 69
 QY 94 TI-----GSILRELDTSMDGTPEPPQNPVPLG---LQNEVPPQDPVPLEA-LSSRYLG 143
 DB 70 SLIPKGYGRVVEIINILVKQVSAPAAPVPTICPTTTTPPPPPSPAHVPPPY-- 127
 QY 144 DSGLDLDFLDITSAVEKAPAPPEPHNLFCAPGSWENELDHIMEIILGS 196
 DB 128 -----VEPTTQCFPIILHPPGAP-----SAHRPWQMKDLQAIKQEVSSS 166

RESULT 3

US-07-672-483-4
 ; Sequence 4, Application US/07672483
 ; Patent No. 5359029
 ; GENERAL INFORMATION:
 ; APPLICANT: LACROIX, Martial
 ; APPLICANT: ZREIN, Maan
 ; TITLE OF INVENTION: PEPTIDES AND ANALOGUES AND MIXTURES
 ; TITLE OF INVENTION: THEREOF FOR DETECTING ANTIBODIES TO HTLV-I AND HTLV-II
 ; TITLE OF INVENTION: VIRUSES
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: FISH & NEAVE
 ; STREET: 875 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-6250
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/672,483
 ; FILING DATE: 19910302
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/554,258
 ; FILING DATE: 18-JUL-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: HALEY JR, James F
 ; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: IAF8 CIP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212)715-0600
 ; TELEFAX: (212)715-0674
 ; TELEX: 14-8367
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 433 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-672-483-4
 Query Match 9.0%; Score 93; DB 1; Length 433;
 Best Local Similarity 23.1%; Pred. No. 0.89;
 Matches 40; Conservative 24; Mismatches 67; Indels 42; Gaps 7;
 QY 47 PRAPSLRRHVLHNTLQQLAALRLAPAPA-----LPPEPLFLGEEDFSLSA 93
 DB 13 PKAP---RGLSTHHLNLFQAAYRLQRPSPDFQQLRRFLKALKTPILNLPIDYSLA 69
 QY 94 TI-----GSILRELDTSMDGTPEPPQNPVPLG---LQNEVPPQDPVPLEA-LSSRYLG 143
 DB 70 SLIPKGYGRVVEIINILVKQVSAPAAPVPTICPTTTTPPPPPSPAHVPPPY-- 127
 QY 144 DSGLDLDFLDITSAVEKAPAPPEPHNLFCAPGSWENELDHIMEIILGS 196
 DB 128 -----VEPTTQCFPIILHPPGAP-----SAHRPWQMKDLQAIKQEVSSS 166
 RESULT 4
 US-09-413-814-28
 ; Sequence 28, Application US/09413814
 ; Patent No. 6225064
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
 ; APPLICANT: Bristol-Myers Squibb, Co.
 ; APPLICANT: Beyer, Stefan
 ; APPLICANT: Bloeker, Helmut
 ; APPLICANT: Brandt, Petra
 ; APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
TITLE OF INVENTION: heteropolyketide compounds
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 28
LENGTH: 2618
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-28

Query Match 9.0%; Score 93; DB 3; Length 2618;
Best Local Similarity 25.5%; Pred. No. 9.2;
Matches 55; Conservative 18; Mismatches .79; Indels 64; Gaps 11;

QY 31 QALLR---ISLD-----KQVSLGPRAP---SLRRHVLHNTLQQLOA 67
DB 974 QALLRRAIGLDERPFGAGNSFGLIRLHAKLESAPGKSFPTDLPQHTSIRSOAEMLSG 1033

QY 68 ALRLAP-APALPPEPLFGEEDFSLSATIGSLRELDTSMDGTEPPQ---NPVTPGLQGN 123
DB 1034 SSVEAPLAGNPQPPAAQAQVASSAKSPGEGAATSSGLTAQPPQPHFRPIAVIGLAG 1093

QY 124 EYVPOPD-PVFLEAL-----SSRYLGDGSLD-----DFFLD 153
DB 1094 RPPAAPDLDAFLELLTEGRCGIRFPFQAELRDEGLDANRIACHNVPAKGFLLDRADHF-D 1152

QY 154 IDTSAVEKEPAR-APPEPHNLFCAPGSWEWNELDH 188
DB 1153 ADFFGIPPRDAEITDQIRLLLECC-----WNALSH 1183

RESULT 5
PCT-US95-04910-13
Sequence 13, Application PC/TUS9504910
GENERAL INFORMATION:
APPLICANT: The Government of the United
APPLICANT: States of America as represented
APPLICANT: by the Secretary, Department of
APPLICANT: Health and Human Services
TITLE OF INVENTION: ISOLATION AND
TITLE OF INVENTION: CHARACTERIZATION OF A NOVEL PRIMATE T-CELL
TITLE OF INVENTION: LYMPHOTROPIC VIRUS AND THE USE OF THIS VIRUS
TITLE OF INVENTION: OR COMPONENTS THEREOF IN DIAGNOSTIC ASSAYS
TITLE OF INVENTION: AND VACCINES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
FILING DATE: 21-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04910
FILING DATE: 22-APR-1994

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4125PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acids
STRANDEDNESS: unknown
TOPOLOGY: unknown
PCT-US95-04910-13

Query Match 9.0%; Score 92.5; DB 5; Length 432;
Best Local Similarity 21.5%; Pred. No. 0.99;
Matches 40; Conservative 19; Mismatches 19; Indels 79; Gaps 7;

QY 47 PRAPSLRRHVLHNTLQLOAALRIAPAPA-----LPPPEPLFGEEDFSLSA 93
DB 13 PKAP-----RGLSTHWNLFQASYRLQPGFSDFFQQLRRFLKALKTPIWLNPIIDYSLIA 69

QY 94 TI-----GSILRELDTSMDGTEPPQNPVTPGLQNEVPPQDPVFLSALSSRYLGD 144
DB 70 SLIPKGYPGRTSEINVLIRNOAGSTPPPPAPSLP-----BPANPPPL----- 111

QY 145 SGLDDFFLDIDTSAVEKEPARAPPEP-----PHNLFCAPGSWEWNE 185
DB 112 -----QQPS-APPEHTPPPYIEPPATHCLPILPHGAPSAHRPAQMKD 154

QY 186 LDHIME 191
DB 155 LQAIKQ 160

RESULT 6

US-09-949-016-10257
Sequence 10257, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10257
LENGTH: 379
TYPE: PRT
ORGANISM: Human
US-09-949-016-10257

Query Match 8.9%; Score 92; DB 4; Length 379;
Best Local Similarity 21.6%; Pred. No. 0.93;
Matches 44; Conservative 23; Mismatches 23; Indels 82; Gaps 10;

QY 3 GGLKRXKHSDLSEEEER-----NEWSPAGLSQSQQALLRLSLDKVQSLGPR 48
DB 221 GGAAPDEDDDDDELEVLRLVETPGRELWVPAAG-----RAARGQAERAGQPS 271

QY 49 APSLRRHVLHNTLQLOAALRIAPAPA-----LPPPEPLFGEEDFSLSATIGSLRELDTSMDG 108

```
Db 272 G-----EGAAAAAASPTPSEDEP--EEEEEEAAAABEG-----EETVASG 312
QY 109 TEPPQNPVPLGLQNEVPPQDPVLEALSSRYLGDGLDDFFLDITSAVEKEPARAPP 168
Db 313 EE-----SLGFLSRLEP-----GPAGL-----DCSALDRDPDLQPP 343
QY 169 EPPHNL-----FCAP-----GSW 181
Db 344 SGTSHFEFPDYCTPEVTMIAGDW 367

RESULT 7
US-09-377-285B-16
; Sequence 16, Application US/09377285B
; Patent No. 6720175
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: WORLEY, Paul
; APPLICANT: TU, Jian
; APPLICANT: XIAO, Bo
; APPLICANT: LEAHY, Daniel
; APPLICANT: BENEKEN, Jutta
; APPLICANT: LANAHAN, Anthony
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING HOMER 1b PROTEIN (AS AMENDED)
; FILE REFERENCE: JHU1580-4
; CURRENT APPLICATION NUMBER: US/09/377,285B
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/138,426
; PRIOR FILING DATE: 1999-06-10
; PRIOR APPLICATION NUMBER: US 60/138,493
; PRIOR FILING DATE: 1999-06-10
; PRIOR APPLICATION NUMBER: US 60/138,494
; PRIOR FILING DATE: 1999-06-10
; PRIOR APPLICATION NUMBER: US 60/097,334
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 16
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-377-285B-16

Query Match 8.8%; Score 90.5; DB 4; Length 370;
Best Local Similarity 23.5%; Pred. No. 1.3;
Matches 43; Conservative 29; Mismatches 62; Indels 49; Gaps 8;

QY 33 LLRLSDKV-QRSLGFRAPSLRRHVLINHTLQQLAALRLAPALPP-----EPLFLG 85
Db 102 MVNMLEKVARREICTLA-----TVRLPPSQKVTPPESLPLTPYCRKPL--- 147
QY 86 EEDFSLSATIGSILRELDTSMDGT-----EPPQNPVPLGLQNEVPPQ-PDPVFLEAL-- 137
Db 148 --NFACLDDVGHGVKDLSTQLSRKTSIKAPATPASATLGRPPRIPEPVQLPAVPD 205
QY 138 -----SSRYLGDGLDDFFLDITSAVEKEPARAPPEP-----PHNLFCAFG 179
Db 206 GKLSASVSSLASAGSAGSAGSIPQSKGVAVATPPPPPIATVTPPPPLPAEFLPLP 265
QY 180 SWE 182
Db 266 PME 268

RESULT 8
US-08-887-518-2
; Sequence 2, Application US/08887518
; Patent No. 5843721
; GENERAL INFORMATION:
; APPLICANT: Rothe, Mike
; APPLICANT: Wu, Lin
; TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
; NUMBER OF SEQUENCES: 4
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,518
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: T97-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 947 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-887-518-2

Query Match 8.8%; Score 90.5; DB 2; Length 947;
Best Local Similarity 25.8%; Pred. No. 4.3;
Matches 54; Conservative 17; Mismatches 87; Indels 51; Gaps 10;

QY 3 GGLKRKHSDLDEEERWESWSPAGLQSYQALLRLISLKVQSLGFRAPSLRRHVLINHTL 62
Db 659 GGLK---SPWRGEVKEPRHPPPNQANYHQ-----TLHAQPRELSRAPGPRPAEETTCRA 710
QY 63 QQLQAALRLAPALPPE-----PLFLGEEDFSLSATIGSILRELDTSMDGTPEPPQNPVT 117
Db 711 PKLQ-----PPLPPEPPPEPNKSPPLTLSEKE-----SGMWEPLFLSLLEBPAPARNPSS 758
QY 118 PLGLQNEVPPQ-----PDPVFLEALSSRY-----LGDGLDDFFLDITSAVEKEPA 164
Db 759 P-ERKATVPBELQQLLELFLNLSQPFSLSEEQEQLISLISLSDDS---EKNPS 814
QY 165 RAPPEPPHNLFCAPGSW-----EWN 184
Db 815 KASQSSRDTLSSGVHWSWSSQAERSSWN 843

RESULT 9
US-09-023-321-2
; Sequence 2, Application US/09023321
; Patent No. 5844073
; GENERAL INFORMATION:
; APPLICANT: Rothe, Mike
; APPLICANT: Wu, Lin
; TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/887,518
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 947 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-321-2

Query Match 8.8%; Score 90.5; DB 2; Length 947;
Best Local Similarity 25.8%; Pred. No. 4.3;
Matches 54; Conservative 17; Mismatches 87; Indels 51; Gaps 10;

QY 3 GGLKREKHSDEEEERWESWSPAGLOSQOALLRISLDKQVRSIGPRLRRHVLHNTL 62
DB 659 GGLK---SPWRGEYKEPRHPPPNQANYHQ-----TLHAQPRELSPRAPGPRPAETTGRA 710
QY 63 QQLQAALRLAPALPPE-----PLFLGEEFSLSATIGSILRELDTSMDGTPEPPQNPVT 117
DB 711 PKLQ-----PPLPPEPPENKSPPLTLSKEE-----SGMWEPLPLSSLEPAPARNPSS 758
QY 118 PLGLQNEVPPQ-----PDVFLFLEALSSRY-----LGDSDLDDFFLDIDTSAVEKEPA 164
DB 759 P-ERKATVPEQLQLEIFLNSLSQPFSEBQILSCLSIDLSLSDDS---EKNPS 814
QY 165 RAPPEPPHNLFCAPGSW-----EWN 184
DB 815 KASQSRDRTLSSGVHSSWSSQAEARSSWN 843

RESULT 10
US-09-032-475-2
Sequence 2, Application US/09032475
Patent No. 5854003
GENERAL INFORMATION:
APPLICANT: Rothe, Mike
APPLICANT: Wu, Lin
TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,475
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,518
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 947 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-032-475-2

Query Match 8.8%; Score 90.5; DB 2; Length 947;
Best Local Similarity 25.8%; Pred. No. 4.3;
Matches 54; Conservative 17; Mismatches 87; Indels 51; Gaps 10;

QY 3 GGLKREKHSDEEEERWESWSPAGLOSQOALLRISLDKQVRSIGPRLRRHVLHNTL 62
DB 659 GGLK---SPWRGEYKEPRHPPPNQANYHQ-----TLHAQPRELSPRAPGPRPAETTGRA 710
QY 63 QQLQAALRLAPALPPE-----PLFLGEEFSLSATIGSILRELDTSMDGTPEPPQNPVT 117
DB 711 PKLQ-----PPLPPEPPENKSPPLTLSKEE-----SGMWEPLPLSSLEPAPARNPSS 758
QY 118 PLGLQNEVPPQ-----PDVFLFLEALSSRY-----LGDSDLDDFFLDIDTSAVEKEPA 164
DB 759 P-ERKATVPEQLQLEIFLNSLSQPFSEBQILSCLSIDLSLSDDS---EKNPS 814
QY 165 RAPPEPPHNLFCAPGSW-----EWN 184
DB 815 KASQSRDRTLSSGVHSSWSSQAEARSSWN 843

RESULT 11
US-09-257-703-1
Sequence 1, Application US/09257703
Patent No. 6265538
GENERAL INFORMATION:
APPLICANT: Greene, Warner C.
APPLICANT: Lin, Xin
APPLICANT: Gelezuinas, Romas
TITLE OF INVENTION: A NOVEL INHIBITOR OF THE INFLAMMATORY RESPONSE INDUCED
TITLE OF INVENTION: BY TNF-ALPHA AND IL-1
FILE REFERENCE: 30448, 61USU1
CURRENT APPLICATION NUMBER: US/09/257,703
CURRENT FILING DATE: 1999-02-25
EARLIER APPLICATION NUMBER: 60/076,299
EARLIER FILING DATE: 1998-02-27
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 947
TYPE: PRT
ORGANISM: NP-KAPPA B INDUCING KINASE (NIK)
US-09-257-703-1

Query Match 8.8%; Score 90.5; DB 3; Length 947;
Best Local Similarity 25.8%; Pred. No. 4.3;
Matches 54; Conservative 17; Mismatches 87; Indels 51; Gaps 10;

QY 3 GGLKREKHSDEEEERWESWSPAGLOSQOALLRISLDKQVRSIGPRLRRHVLHNTL 62
DB 659 GGLK---SPWRGEYKEPRHPPPNQANYHQ-----TLHAQPRELSPRAPGPRPAETTGRA 710
QY 63 QQLQAALRLAPALPPE-----PLFLGEEFSLSATIGSILRELDTSMDGTPEPPQNPVT 117
DB 711 PKLQ-----PPLPPEPPENKSPPLTLSKEE-----SGMWEPLPLSSLEPAPARNPSS 758
QY 118 PLGLQNEVPPQ-----PDVFLFLEALSSRY-----LGDSDLDDFFLDIDTSAVEKEPA 164

Db 759 P-ERKATVPEQLQLEIEFLNLSQPFSLSEEQILSCLSIDSLSDDS---EKNPS 814
QY 165 RAPPEPHNLCAPGSW-----EWN 184
Db 815 KASQSSRDTLSSGVHSSWSQAEARSSWN 843

RESULT 12

US-09-871-889A-1
; Sequence 1, Application US/09871889A
; Patent No. 6645728
; GENERAL INFORMATION:
; APPLICANT: Greene, Warner C.
; APPLICANT: Lin, Kin
; APPLICANT: Gelezinas, Romas
; TITLE OF INVENTION: A NOVEL INHIBITOR OF THE INFLAMMATORY RESPONSE INDUCED BY TNF- α
; FILE REFERENCE: 30448.61USD1
; CURRENT APPLICATION NUMBER: US/09/871,889A
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 09/257,703
; PRIOR FILING DATE: 1999-02-25
; PRIOR APPLICATION NUMBER: 60/076,299
; PRIOR FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 947
; TYPE: PRT
; ORGANISM: NF-KAPPA B INDUCING KINASE (NIK)
US-09-871-889A-1

Query Match 8.8%; Score 90.5; DB 4; Length 947;
Best Local Similarity 25.8%; Pred. No. 4.3;
Matches 54; Conservative 17; Mismatches 87; Indels 51; Gaps 10;
QY 3 GGLKKHSDLEEBEREWSPAGLSQYQALLRISLQVQSLGPRAPSLRRHVLHNTL 62
Db 659 GGLK---SPWGEYKEPRHPPPNQANYHQ-----TLHAQPRELSPRAGPRAETTGRA 710
QY 63 QOLQAALRLAPALAPPE-----PLFLGEEDFSLSATIGSILRELDTSMDGTPEPPQNPT 117
Db 711 PKLQ-----PPLPPEPRKNSPPLTSKEE-----SGMWEPLPLSSLEPAPARNPSS 758
QY 118 PLGLQNEVPPQ-----PDVFLHLSRY-----LQDGLDDFFLDIDTSAVEKEPA 164
Db 759 P-ERKATVPEQLQLEIEFLNLSQPFSLSEEQILSCLSIDSLSDDS---EKNPS 814
QY 165 RAPPEPHNLCAPGSW-----EWN 184
Db 815 KASQSSRDTLSSGVHSSWSQAEARSSWN 843

RESULT 13

US-09-902-540-11750
; Sequence 11750, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11750
; LENGTH: 1027
; TYPE: PRT

; ORGANISM: Myxococcus xanthus
US-09-902-540-11750
Query Match 8.8%; Score 90.5; DB 4; Length 1027;
Best Local Similarity 25.3%; Pred. No. 4.7;
Matches 42; Conservative 12; Mismatches 55; Indels 57; Gaps 7;
QY 36 ISLDKVSRLGPRAPSLRRHVLHNTLQOLQAALRLAPALAPPEPLFLGEEDFSLSATI 95
Db 271 LDLSDAEAAPPRAP-----VAPKAPPAPV-----TSPAGS 301
QY 96 GSILRELDTSMDGTPEPPQNPTPLGLQNEVPPQDPV-----FLEALSSR----- 140
Db 302 GGIEFDLDAEDVSPPP--PVAPPPPRAPAAPPPAPVSGVGFEDFALSDDAGTVPPPP 359
QY 141 -----VLGSGLDFFLDIDTSA-VEKEPARAPPEPPHNLFCAP 178
Db 360 PRAPSPVSAGGGID---FDLGLSEDVEEAPVVPVPPPPRAAMVPP 402

RESULT 14

US-08-778-717-9
; Sequence 9, Application US/08778717
; Patent No. 6602689
; GENERAL INFORMATION:
; APPLICANT: UENO, EIICHI
; APPLICANT: NOBUYUKI, FUJII
; APPLICANT: OKADA, MASAHISA
; TITLE OF INVENTION: FUSED DNA SEQUENCE, FUSED PROTEIN
; TITLE OF INVENTION: EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR
; TITLE OF INVENTION: EXPRESSING SAID DNA SEQUENCE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/778,717
; FILING DATE: 12-DEC-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER: JP 352225/1995
; FILING DATE: 28-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2084-031-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: RECOMBINANT
; PUBLICATION INFORMATION:
; AUTHORS: NOBUYUKI FUJII ET AL.
; TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED
; TITLE: FROM SAID FUSED DNA SEQUENCE AND METHOD OF
; TITLE: EXPRESSING SAID FUSED PROTEIN

```
; RELEVANT RESIDUES IN SEQ ID NO: 9: FROM 1 TO 215
US-08-778-717-9

Query Match      8.7%; Score 89.5; DB 4; Length 215;
Best Local Similarity 23.7%; Pred. No. 0.78;
Matches 41; Conservative 24; Mismatches 67; Indels 41; Gaps 6;

QY 2 EGGLKHKHSDLEERE-----ERWWSFAGLSQYQALLRISLDKQVRSLG-----46
Db 31 EGKAQPKVEVEDESLYNCAKEAMEACPVSAITIEEAGGSSLVPRGSFQMGIHGLSPTP 90
QY 47 -PRAPSLRRHVLHNTLQOALRLAPAPA-----LPPPEFLFGEEDFSL 92
Db 91 IPKAP--RGLSTHWLFLQAYLQRPSPDFDQQLRRFLKALKTPILWLPIDYSL 147
QY 93 ATI-----GSTILRELDTSMDGTEPPQNPVTPLG---LQNEVPPQDPVFLEA 136
Db 148 ASLIPKGYGRVVEIINILVKQVSFSAAPAVPTPICTTTTPPPPPPSPEA 200

RESULT 15
US-09-949-016-7139
; Sequence 7139, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7139
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7139

Query Match      8.7%; Score 89.5; DB 4; Length 611;
Best Local Similarity 25.8%; Pred. No. 3;
Matches 48; Conservative 21; Mismatches 56; Indels 61; Gaps 9;

QY 12 LEEEREWESFAGLSQYQALLRISLDKQVRSFGRAPSLRRHVLHNTLQOALRL 71
Db 435 LQKEQQR---SKLESQR-----SLEQANRSLQLRIQELQLQAIH-----GLPV 477
QY 72 APAP-----ALPPEPLFGEEDFSLSATIGSILRELDTSMDGTEPPQNPVTPL 119
Db 478 PPTGLLSLATTSASDSLKPSQDIEEGRPGATFHV-----GGGPAQN-----522
QY 120 GLQNEVPPQP--DPVFLEALSSRYLGDGLDFFLDIDTSAVEKE-----PA 164
Db 523 -APHQOPAPPSDALLDLHFFPSDHLGLG-DPFLHGLEDLMEEEGVVGLSGGALSPL 580
QY 165 RAPPEP 170
Db 581 RAASDP 586
```

Search completed: April 25, 2005, 12:34:02
Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2005, 03:22:52 ; Search time 2749 Seconds
(without alignments)
10417.254 Million cell updates/sec

Title: US-10-069-386A-1
Perfect score: 591
Sequence: 1 atggaggaggctgaagag.....aaatcattctgggtcctaa 591

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	591	100.0	591	6	AX088925 Sequence
2	591	100.0	1542	9	AF192529 Homo sapi
3	589.4	99.7	1325	6	CQ783944 Sequence
4	589.4	99.7	1325	6	BD127913 Primer fo
5	589.4	99.7	1325	9	AK074604 Homo sapi
6	589.4	99.7	1346	9	BC014061 Homo sapi
7	589.4	99.7	1420	9	BC050643 Homo sapi
8	589.4	99.7	1536	6	CQ731553 Sequence
9	589.4	99.7	160643	9	AC010271 Homo sapi
10	588.4	99.6	825	6	CQ779981 Sequence
11	588.4	99.6	825	6	CQ781937 Sequence
12	588.4	99.6	825	6	BD124690 Primer fo
13	588.4	99.6	825	6	BD126646 Primer fo
14	584.6	98.9	164652	2	AC021625 Homo sapi
15	473	80.0	506	6	CQ463286 Sequence
16	429.2	72.6	256418	2	AC115309 Rattus no
17	429.2	72.6	257945	2	AC112313 Rattus no
18	429.2	72.6	290929	2	AC118914 Rattus no
19	426	72.1	657	10	AF317202 Mus muscu

20	426	72.1	1356	10	BC034886	BC034886 Mus muscu
21	426	72.1	145347	2	AC073821	AC073821 Mus muscu
22	426	72.1	201377	2	AC073767	AC073767 Mus muscu
23	186.6	31.6	205	6	CQ463461	CQ463461 Sequence
24	54.6	9.2	148851	9	HS155D22	297205 Human DNA s
25	49.4	8.4	68050	2	AC079074	AC079074 Homo sapi
26	49.2	8.3	7218	6	I66494	166494 Sequence 14
27	48.6	8.2	125020	9	AF429315	AF429315 Homo sapi
C 28	48	8.1	92564	9	AY007685	AY007685 Homo sapi
C 29	48	8.1	125020	9	AF429315	AF429315 Homo sapi
30	47.4	8.0	2082	10	AF322238	AF322238 Mus muscu
31	47.4	8.0	2432	10	BC055824	BC055824 Mus muscu
32	47.4	8.0	163295	2	AC102702	AC102702 Mus muscu
33	47.4	8.0	167753	2	AC126039	AC126039 Mus muscu
C 34	46.4	7.9	3179	6	AX003139	AX003139 Sequence
C 35	46.4	7.9	25138	9	HSTERT2	AF128894 Homo sapi
C 36	46.4	7.9	51552	6	AR266023	AR266023 Sequence
37	46.4	7.9	202305	9	AC114291	AC114291 Homo sapi
C 38	46	7.8	173910	9	AC083982	AC083982 Homo sapi
C 39	46	7.8	302176	2	AC096479	AC096479 Rattus no
40	45.6	7.7	37405	2	AC101273	AC101273 Mus muscu
41	45.6	7.7	148213	10	AL645731	AL645731 Mouse DNA
42	45.6	7.7	189378	2	AC148826	AC148826 Pan trogl
43	45.4	7.7	189230	2	AC120836	AC120836 Mus muscu
C 44	45.2	7.6	136753	2	AC149968	AC149968 Strongylo
45	45	7.6	69429	2	AC100015	AC100015 Mus muscu

ALIGNMENTS

RESULT 1	AX088925	Sequence 1 from Patent WO0114546.	591 bp	DNA	linear	PAT 17-MAR-2001
LOCUS	AX088925	Sequence 1 from Patent WO0114546.	591 bp	DNA	linear	PAT 17-MAR-2001
DEFINITION	AX088925	Sequence 1 from Patent WO0114546.	591 bp	DNA	linear	PAT 17-MAR-2001
ACCESSION	AX088925	Sequence 1 from Patent WO0114546.	591 bp	DNA	linear	PAT 17-MAR-2001
VERSION	AX088925.1	GI:13397684	591 bp	DNA	linear	PAT 17-MAR-2001
KEYWORDS	unidentified					
SOURCE	unidentified					
ORGANISM	unclassified.					
REFERENCE	1	Alaoui-Jamali, M.A. and Cho, J.M.				
AUTHORS	Alaoui-Jamali, M.A. and Cho, J.M.					
TITLE	Replication protein A binding transcriptional factor (rbt1) and uses thereof					
JOURNAL	Patent: WO 0114546-A 1 01-MAR-2001;					
FEATURES	Centre for Translational Research in Cancer (CA)					
source	Location/Qualifiers					
	1..591					
	/organism="unidentified"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:32644"					
	/note="replication protein A transcriptional factor"					

ORIGIN

Query Match	100.0%	Score 591;	DB 6;	Length 591;
Best Local Similarity	100.0%	Pred. No. 7.3e-120;		
Matches 591;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGGAGGAGGCTTCAAGAGGAAACACTCTGATTTGGAGAGGAGGAGAGGTGGGAG	60	
DB	1	ATGGAGGAGGCTTCAAGAGGAAACACTCTGATTTGGAGAGGAGGAGAGGTGGGAG	60	
QY	61	TGGAGTCCAGAGGCTTCCAGAGCTACACAGAGCCCTGCTCCGATCTCCCTAGACAA	120	
DB	61	TGGAGTCCAGAGGCTTCCAGAGCTACACAGAGCCCTGCTCCGATCTCCCTAGACAA	120	
QY	121	GTCCAGCGAGCTCGGGCCCCCGAGCACCCAGCCCTCCGAGGCATGTCTCATCCATAAC	180	
DB	121	GTCCAGCGAGCTCGGGCCCCCGAGCACCCAGCCCTCCGAGGCATGTCTCATCCATAAC	180	
QY	181	ACCTTCCAGAGCTCGAGGCTGTCATCTGGCTGGCTCCGCCCTGCTCCCTGCCCCCGAG	240	
DB	181	ACCTTCCAGAGCTCGAGGCTGTCATCTGGCTGGCTCCGCCCTGCTCCCTGCCCCCGAG	240	

[illegible]

KEPARAPPEPPHNLFCAPGSWEWNELDHTMETILGS"

ORIGIN

Query Match	99.7%;	Score 589.4;	DB 6;	Length 1325;
Best Local Similarity	99.8%;	Pred. No. 1.5e-119;		
Matches 590;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	ATGAGGAGGAGCTTGAAGAGGAAACACTCTGATTTTGGAAAGAGGAGGAGGAGGTGGGAG	60	
DB	93	ATGGTGGAGGCTTGAAGAGGAAACACTCTGATTTTGGAAAGAGGAGGAGGAGGTGGGAG	152	
QY	61	TGGAGTCCAGCAGGCCCTTCAGAGCTACGACGAGCCCTGTCTCCGATCTCTCCCTAGACAAA	120	
DB	153	TGGAGTCCAGCAGGCCCTTCAGAGCTACGACGAGCCCTGTCTCCGATCTCTCCCTAGACAAA	212	
QY	121	GTCCAGGCGAGCCTGGGCCCCCGAGCACCAGCCCTCCGCGAGGATGCTCTCATCCATAAC	180	
DB	213	GTCCAGGCGAGCCTGGGCCCCCGAGCACCAGCCCTCCGCGAGGATGCTCTCATCCATAAC	272	
QY	181	ACCTCCAAACAGCTGCAGGGTGCACCTTCGCTTGCTCCCGCCCTGCCTGCCCCCGGAG	240	
DB	273	ACCTCCAAACAGCTGCAGGGTGCACCTTCGCTTGCTCCCGCCCTGCCTGCCCCCGGAG	332	
QY	241	CCCTCTTCTTGGGCGAGGAGATTTCTCCCTGTTCAGGCACCATTTGGGTCTATCCTCAGG	300	
DB	333	CCCTCTTCTTGGGCGAGGAGATTTCTCCCTGTTCAGGCACCATTTGGGTCTATCCTCAGG	392	
QY	301	GAGCTGACACCTCCATGGATGGGATGAGCCCTCAGATCCAGTGACTCCCTCTGGC	360	
DB	393	GAGCTGACACCTCCATGGATGGGATGAGCCCTCAGATCCAGTGACTCCCTCTGGC	452	
QY	361	CTCCAGAAATGAAGTGCCACCCAGCCCTGATCCAGTCTTTCTTAGAGCTCTGAGCTCCCGG	420	
DB	453	CTCCAGAAATGAAGTGCCACCCAGCCCTGATCCAGTCTTTCTTAGAGCTCTGAGCTCCCGG	512	
QY	421	TACTTTGGGGACTCTGGCGCTGGATGACTTTCTTTCTGGACATTGACATCTGCGGTGAA	480	
DB	513	TACTTTGGGGACTCTGGCGCTGGATGACTTTCTTTCTGGACATTGACATCTGCGGTGAA	572	
QY	481	AAGGAGCCTGCACGGGGCCACACAGAGCCTCCTCAACACCTCTTCTGTGCCCCAGGTTCT	540	
DB	573	AAGGAGCCTGCACGGGGCCACACAGAGCCTCCTCAACACCTCTTCTGTGCCCCAGGTTCT	632	
QY	541	TGGGAGTGGAAATGAATGGATTCACATCATGGAATCAATTCCTGGGGTCTCTAA	591	
DB	633	TGGGAGTGGAAATGAATGGATTCACATCATGGAATCAATTCCTGGGGTCTCTAA	683	

RESULT 4	
BD127913	1325 bp DNA linear PAT 18-SEP-2002
LOCUS	
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.
ACCESSION	
VERSION	BD127913.1 GI:23222858
KEYWORDS	JP 2002017375-A/3344.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1. (bases 1 to 1325)
AUTHORS	Oca, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y., Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and Koga, H.
TITLE	Primer for synthesizing full-length cDNA and use thereof
JOURNAL	Patent: JP 2002017375-A 3344 22-JAN-2002;
	HELIX RESEARCH INSTITUTE
COMMENT	OS Homo sapiens (human)
	PN JP 2002017375-A/3344
	PD 22-JAN-2002
	PF 07-JUL-2000 JP 2000253172
	PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
	ISHII,
	PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI

```

SHINICHI KOJIMA.
PI TETSUJI OTSUKI, HISASHI KOGA
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH key
Location/Qualifiers
FT CDS Location/Qualifiers
1..1325
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
FEATURES
source

```

ORIGIN

source

ORIGIN

Query Match	99.7%	Score 589.4	DB 6	Length 1325
Best Local Similarity	99.8%	Pred. No. 1.5e-119		
Matches 590	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1	ATCGAGGGAGCCTTGAAGAGGAAACACTCTGTATTTGGAAGAGGAGGAGGAGGTGGGAG	60	
DB	93	ATCGTGGGAGCCTTGAAGAGGAAACACTCTGTATTTGGAAGAGGAGGAGGAGGTGGGAG	152	
QY	61	TGGAGTCCAGCAGGCGCTTCAGAGCTACCAAGAACCCCTGTCTCGCATCTCCCTAGACAAA	120	
DB	153	TGGAGTCCAGCAGGCGCTTCAGAGCTACCAAGAACCCCTGTCTCGCATCTCCCTAGACAAA	212	
QY	121	GTCCAGGGGAGCTGGGCCCCCGAGCACACAGAGCTCCGCAAGGCATGTCTCATCCATAAC	180	
DB	213	GTCCAGGGGAGCTGGGCCCCCGAGCACACAGAGCTCCGCAAGGCATGTCTCATCCATAAC	272	
QY	181	ACCCTCCAAAGCTGCAGGCTGCACCTTCGGCTTGCGCTCCCGCCCCGTGCCCTCCCGCGAG	240	
DB	273	ACCCTCCAAAGCTGCAGGCTGCACCTTCGGCTTGCGCTCCCGCCCCGTGCCCTCCCGCGAG	332	
QY	241	CCGCTCTTCTGGGCGAGGAGGATTTCTCCCTGTACGCAACCATTTGGCTCTATCTCAGG	300	
DB	333	CCGCTCTTCTGGGCGAGGAGGATTTCTCCCTGTACGCAACCATTTGGCTCTATCTCAGG	392	
QY	301	GAGCTGACACCTTCATGGATGGAGCTGAGCCCCCTCAGAAATCCAGTGCCTCCCTTGGC	360	
DB	393	GAGCTGACACCTTCATGGATGGAGCTGAGCCCCCTCAGAAATCCAGTGCCTCCCTTGGC	452	
QY	361	CTCCAGAAATGAAGTGGCAACCCAGCGCTGATCCAGTCTTCTTTAGAACTCTGAGCTCCCGG	420	
DB	453	CTCCAGAAATGAAGTGGCAACCCAGCGCTGATCCAGTCTTCTTTAGAACTCTGAGCTCCCGG	512	
QY	421	TACTTTGGGAGCTCTGGCGTGGATGACTTCTTTCTTGGACATTTGACACATCTGCGGTAGAA	480	
DB	513	TACTTTGGGAGCTCTGGCGTGGATGACTTCTTTCTTGGACATTTGACACATCTGCGGTAGAA	572	
QY	481	AAGGAGCCTGCAAGGGCCCCCAGAGGCTCTCTCACAACTCTTCTGTGCCCCAGGTTCT	540	
DB	573	AAGGAGCCTGCAAGGGCCCCCAGAGGCTCTCTCACAACTCTTCTGTGCCCCAGGTTCT	632	
QY	541	TGGGAGTGGAAATGAATGGATCAATCATCGAAATCAATCTCGGGGTCTCTAA	591	
DB	633	TGGGAGTGGAAATGAATGGATCAATCATCGAAATCAATCTCGGGGTCTCTAA	683	

RESULTS

RESULI 3
AK074604

AKU/4604
1.0015

LOCUS
DEFINITION

DEFINITION

Abstract:

ACCESSION
NUMBER

VERSION

KEYWORDS

SOURCE

ORGANI

AK074604 1325 bp mRNA linear PRI 03-SEP-2008
Homo sapiens cDNA FLJ90123 fis, clone HEMBA1007226, highly similar
to Homo sapiens RPA-binding trans-activator (RBT1) mRNA.
AK074604
AK074604.1 GI:22760150
oligo capping; fis (full insert sequence).
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CDs	/note="synonym: RBT1"	
	/db_xref="LocusID:29946"	
	168_758	
	/gene="SERTAD3"	
	/codon_start=1	
	/product="RPA-binding trans-activator"	
	/protein_id="AAH50643.1"	
	/db_xref="GI:29791865"	
	/db_xref="LocusID:29946"	
	/translation="MVGSLKRLKHSLEBERWSPAGLSYQALLRLISLDKVKQRS	
	LGPAPSLRRHLVHINTLQQLAALPAPALPPEPLFLGEEEDPSLSATIGSLREL	
	DTSMDGTPPQNPVPLGLQNEVPDPVFLFLEALSSRYLGDGLDDFLDITSAVE	
	KEPARAPPEPHNLFCAFGSEWNELDHIMEILGS"	
ORIGIN	Query Match	
	Best Local Similarity 99.7%; Score 589.4; DB 9; Length 1420;	
	Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 ATGGAGGGAGGCTTCAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGAGGTGGGAG 60	
	168 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 227	
QY	61 TGGAGTCCAGCAGGCTTCCAGAGTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 120	
	228 TGGAGTCCAGCAGGCTTCCAGAGTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 287	
QY	121 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCCTCCGAGGCATGTCTCATCCATAAC 180	
	288 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCCTCCGAGGCATGTCTCATCCATAAC 347	
QY	181 ACCCTCCAAACAGCTGAGGCTGACCTTGGCTGGCTCCGCCCTCCGCCCTCCGCCGAG 240	
	348 ACCCTCCAAACAGCTGAGGCTGACCTTGGCTGGCTCCGCCCTCCGCCCTCCGCCGAG 407	
QY	241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300	
	408 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 467	
QY	301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTCACTCCCTTGGC 360	
	468 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTCACTCCCTTGGC 527	
QY	361 CTCCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTCTTGAAGCTCTGAGCTCCCGG 420	
	528 CTCCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTCTTGAAGCTCTGAGCTCCCGG 587	
QY	421 TACTTGGGGGACTCTGGCTGGATGACTTCTTCTGGACATTTGACACATCTGCGGTAGAA 480	
	588 TACTTGGGGGACTCTGGCTGGATGACTTCTTCTGGACATTTGACACATCTGCGGTAGAA 647	
QY	481 AAGGAGCTTCGACGGGCCCCCAGAGCCTCTCTCAACCTTCTTGTGCCCCAGGTTCT 540	
	648 AAGGAGCTTCGACGGGCCCCCAGAGCCTCTCTCAACCTTCTTGTGCCCCAGGTTCT 707	
QY	541 TGGGAGTGGAAATGAATCGATCAGTCAATCGAAATCATTTCTGGGTCCTAA 591	
	708 TGGGAGTGGAAATGAATCGATCAGTCAATCGAAATCATTTCTGGGTCCTAA 758	
RESULT 8	CQ731553	
	Sequence 17487 from Patent WO02068579.	
	CQ731553	
	CQ731553.1 GI:42308440	
KEYWORDS	Homo sapiens (human)	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.	
AUTHORS	JOURNAL	
	Unpublished	
TITLE	Kits, such as nucleic acid arrays, comprising a majority of	
	humanexons or transcripts, for detecting expression and other uses	
	thereof	
	Patent: WO 02068579-A 17487 06-SEP-2002;	
JOURNAL	PE Corporation (US)	
	Location/Qualifiers	
	1. 1536	
	/organism="Homo sapiens"	
FEATURES	/mol_type="unassigned DNA"	
	/db_xref="taxon:9606"	
ORIGIN	Query Match	
	Best Local Similarity 99.8%; Pred. No. 1.4e-119;	
	Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 60	
	290 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 349	
QY	61 TGGAGTCCAGCAGGCTTCCAGAGTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 120	
	350 TGGAGTCCAGCAGGCTTCCAGAGTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 409	
QY	121 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCCTCCGAGGCATGTCTCATCCATAAC 180	
	410 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCCTCCGAGGCATGTCTCATCCATAAC 469	
QY	181 ACCCTCCAAACAGCTGAGGCTGACCTTGGCTGGCTCCGCCCTCCGCCCTCCGCCGAG 240	
	470 ACCCTCCAAACAGCTGAGGCTGACCTTGGCTGGCTCCGCCCTCCGCCCTCCGCCGAG 529	
QY	241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300	
	530 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 589	
QY	301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTCACTCCCTTGGC 360	
	590 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTCACTCCCTTGGC 649	
QY	361 CTCCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTCTTGAAGCTCTGAGCTCCCGG 420	
	650 CTCCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTCTTGAAGCTCTGAGCTCCCGG 709	
QY	421 TACTTGGGGGACTCTGGCTGGATGACTTCTTCTGGACATTTGACACATCTGCGGTAGAA 480	
	710 TACTTGGGGGACTCTGGCTGGATGACTTCTTCTGGACATTTGACACATCTGCGGTAGAA 769	
QY	481 AAGGAGCTTCGACGGGCCCCCAGAGCCTCTCTCAACCTTCTTGTGCCCCAGGTTCT 540	
	770 AAGGAGCTTCGACGGGCCCCCAGAGCCTCTCTCAACCTTCTTGTGCCCCAGGTTCT 829	
QY	541 TGGGAGTGGAAATGAATCGATCAGTCAATCGAAATCATTTCTGGGTCCTAA 591	
	830 TGGGAGTGGAAATGAATCGATCAGTCAATCGAAATCATTTCTGGGTCCTAA 880	
RESULT 9	AC010271	
	Homo sapiens chromosome 19 clone CTC-492K19, complete sequence.	
	AC010271	
	AC010271.8 GI:21743753	
KEYWORDS	HTG.	
	Homo sapiens (human)	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE	1 (bases 1 to 160643)	
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
	DOE Joint Genome Institute and Stanford Human Genome Center.	
	Direct Submission	
JOURNAL	Unpublished	
AUTHORS	2 (bases 1 to 160643)	

AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	3 (bases 1 to 160643)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (29-AUG-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	4 (bases 1 to 160643)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	5 (bases 1 to 160643)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On Jul 13, 2002 this sequence version replaced gi:15887272. Draft sequence produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.sbgc.stanford.edu
FEATURES	Quality: Phrap Quality >=40 100% of Sequence; Estimated Total Number of Errors is 0.1. Location/Qualifiers 1. 160643 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="19" /clone="CTC-492K19"
ORIGIN	
Query Match	99.7%; Score 589.4; DB 9; Length 160643;
Best Local Similarity	99.8%; Pred. No. 7.9e-120;
Matches 590; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 ATGGAGGAGGCTTGAACAGGAAACACTCTGATTTGGAAGAGGAGGAGGAGGTGGGAG 60
Db	137453 ATGTTGGAGGCTTGAAGAGGAAACACTCTGATTTGGAAGAGGAGGAGGAGGTGGGAG 137394
QY	61 TGGAGTCCAGCAGGCTTTTCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 120
Db	137393 TGGAGTCCAGCAGGCTTTTCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 137334
QY	121 GTCAGGCGAGCTTGGGCCCCCGAGCACCCAGCTCCGAGGCGATGCTCATCATTAAC 180
Db	137333 GTCCAGGCGAGCTTGGGCCCCCGAGCACCCAGCTCCGAGGCGATGCTCATCATTAAC 137274
QY	181 ACCCTCCAAAGCTGCAGGCTGCATTCGCTGGCTCCGCCCCCTGCCCTGCCCTCCCGAG 240
Db	137273 ACCCTCCAAAGCTGCAGGCTGCATTCGCTGGCTCCGCCCCCTGCCCTGCCCTCCCGAG 137214
QY	241 CCCCTCTTCTCTGGGCGAGGAGATTTCTCCCTGTTCAGACCACTTCGCTGCTATCTCAGG 300
Db	137213 CCCCTCTTCTCTGGGCGAGGAGATTTCTCCCTGTTCAGACCACTTCGCTGCTATCTCAGG 137154
QY	301 GAGCTGGACACCTCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTGAATCCCTTTGGC 360
Db	137153 GAGCTGGACACCTCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTGAATCCCTTTGGC 137094
QY	361 CTCAGAAATGAAGTGCACCCAGCTCATCAGTCTTCTTAGAGCTCTGAGCTGCCGCGG 420
Db	137093 CTCAGAAATGAAGTGCACCCAGCTCATCAGTCTTCTTAGAGCTCTGAGCTGCCGCGG 137034
QY	421 TACTTGGGGGACTCTGGGCTTGATGACTTCTTTCTGGACATTGACACATCTCGCGTAGAA 480
Db	137033 TACTTGGGGGACTCTGGGCTTGATGACTTCTTTCTGGACATTGACACATCTCGCGTAGAA 136974
QY	481 AAGGAGCCTGCAACGGGCCCCCAACGAGCCTCTCTCAACACCTTCTTGTCGCCCAAGTTCT 540

Db 136973 AAGAGGCTGCACGGGCCCCACAGAGGCTCCTCACACCTCTTCTGTGCCCCAGGTTCT 13691

Qy 541 TGGGAGTGGAAATGAACCTGGATCAATCATCATGGAATCATTTCTGGGGTCTCTAA 591
|||||

Db 136913 TGGGAGTGGAAATGAACCTGGATCAATCATCATGGAATCATTTCTGGGGTCTCTAA 136863
|||||

RESULT 10
CQ779981
LOCUS Homo sapiens (human)
DEFINITION Sequence 121 from Patent EP1396543.
ACCESSION CQ779981
VERSION CQ779981.1 GI:45536053
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
TITLE Primers for synthesizing full length cDNA clones and their use
JOURNAL Patent: EP 1396543-A 121 10-MAR-2004;
RESEARCH Association for Biotechnology (JP)
FEATURES
source Location/Qualifiers
1..825
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 99.6%; Score 588.4; DB 6; Length 825;
Best Local Similarity 99.7%; Pred. No. 2.6e-119;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGAGGGAGGCTTGAAGAGAAACACTCTGATTTGGAAGAGGAGGAGGAGTGGGAG 60
Db 93 ATGCTGGGAGGCTTGAAGAGAAACACTCTGATTTGGAAGAGGAGGAGGAGTGGGAG 152
Qy 61 TGGAGTCCAGCAGGCGCTTCAGAGCTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 120
Db 153 TGGAGTCCAGCAGGCGCTTCAGAGCTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 212
Qy 121 GTCAGGCGAGCCTGGGCCCCCGAGCACCCAGCCTCCGACGCGATGTCTCATCCATAAC 180
Db 213 GTCAGGCGAGCCTGGGCCCCCGAGCACCCAGCCTCCGACGCGATGTCTCATCCATAAC 272
Qy 181 ACCCTCAACAGCTGCAGGCTGCACCTGCTGCTGCTCCGCCCCCTGCCCCCCCCGAG 240
Db 273 ACCCTCAACAGCTGCAGGCTGCACCTGCTGCTGCTCCGCCCCCTGCCCCCCCCGAG 332
Qy 241 CCCTCTCTCTGGGCGAGGAGGATTTCTCCCTGTGACGACCATTTGGCTCTATCTCTCAGG 300
Db 333 CCCTCTCTCTGGGCGAGGAGGATTTCTCCCTGTGACGACCATTTGGCTCTATCTCTCAGG 392
Qy 301 GAGCTGGACACCTCCATGATGGGACTGAGCCCCCTCAGAAATCCAGTGACTCCCCCTTGGC 360
Db 393 GAGCTGGACACCTCCATGATGGGACTGAGCCCCCTCAGAAATCCAGTGACTCCCCCTTGGC 452
Qy 361 CTCAGAAATGAAGTGCCACCCAGCCTGATCCAGTCTTTCTAGAGAGCTGTGAGCTCCCGG 420
Db 453 CTCAGAAATGAAGTGCCACCCAGCCTGATCCAGTCTTTCTAGAGAGCTGTGAGCTCCCGG 512
Qy 421 TACTTTGGGAGCTCTGGCGTGGATGACTTCTTTCTGGACATTTGACACATCTGCGGTAGAA 480
Db 513 TACTTTGGGAGCTCTGGCGTGGATGACTTCTTTCTGGACATTTGACACATCTGCGGTAGAA 572
Qy 481 AAGGAGCTGCACGGGGCCCCACAGAGGCTCCTCACAACTCTTCTGTGCCCCAGGTTCT 540
Db 573 AAGGAGCTGCACGGGGCCCCACAGAGGCTTCTCACAACTCTTCTGTGCCCCAGGTTCT 632
Qy 541 TGGGAGTGGAAATGAACCTGGATCAATCATGGAATCATTTCTGGGGTCTCTAA 591

```
Db 633 TGGGAGTGAATGAATGGATCAGATCATCATGGAATCAATTCGGGTCCTAA 683
|||||
RESULT 11
CQ781937
LOCUS CQ781937 825 bp DNA linear PAT 17-MAR-2004
DEFINITION Sequence 2077 from Patent EP1396543.
ACCESSION CQ781937
VERSION CQ781937.1 GI:45537993
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
TITLE Primers for synthesizing full length cDNA clones and their use
JOURNAL Patent: EP 1396543-A 2077 10-MAR-2004;
RESEARCH Association for Biotechnology (JP)
FEATURES
source
Location/Qualifiers
1..825
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 99.6%; Score 588.4; DB 6; Length 825;
Best Local Similarity 99.7%; Pred. No. 2.6e-119;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 60
|||
DB 93 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 152
|||
QY 61 TGGAGTCCAGCAGGCTTCCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 120
|||
DB 153 TGGAGTCCAGCAGGCTTCCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 212
|||
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCCTCCGAGCATGTCTCATCATCAATAC 180
|||
DB 213 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCCTCCGAGCATGTCTCATCATCAATAC 272
|||
QY 181 ACCCTCCACAGCTGCAGCTGCATTTGGCTTGGCTCCGCCCCCTGCCCCCGAG 240
|||
DB 273 ACCCTCCACAGCTGCAGCTGCATTTGGCTTGGCTCCGCCCCCTGCCCCCGAG 332
|||
QY 241 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTTCAGCCACCATTTGGCTCTATCTCAGG 300
|||
DB 333 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTTCAGCCACCATTTGGCTCTATCTCAGG 392
|||
QY 301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTGACTCCCTTGGC 360
|||
DB 393 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTGACTCCCTTGGC 452
|||
QY 361 CTCGAGATGAAGTGCACCCAGCGCTGATCCAGTCTTCTTAGAGCTCTGAGCTCCCGG 420
|||
DB 453 CTCGAGATGAAGTGCACCCAGCGCTGATCCAGTCTTCTTAGAGCTCTGAGCTCCCGG 512
|||
QY 421 TACTTTGGGGAATCTGGGCTGTGATGACTTCTTCTTGGACATTTGACACATCTGCGGTAGAA 480
|||
DB 513 TACTTTGGGGAATCTGGGCTGTGATGACTTCTTCTTGGACATTTGACACATCTGCGGTAGAA 572
|||
QY 481 AAGAGGCTTGCACGGGCCCCACAGAGGCTCTCTCAACCTTTTGTGCCCCAGGTTCT 540
|||
DB 573 AAGAGGCTTGCACGGGCCCCACAGAGGCTCTCTCAACCTTTTGTGCCCCAGGTTCT 632
|||
QY 541 TGGGAGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 591
|||
DB 633 TGGGAGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 683
|||||
```

```
RESULT 12
BD124690
LOCUS BD124690 825 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD124690
VERSION BD124690.1 GI:23219635
KEYWORDS JP 2002017375-A/121.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 121 22-JAN-2002;
HELIIX RESEARCH INSTITUTE
COMMENT
OS Homo sapiens (human)
PN JP 2002017375-A/121
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI,HISASHI KOGA
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES
FT source
Location/Qualifiers
1..825
/organism="Homo sapiens (human)".
FT source
Location/Qualifiers
1..825
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 99.6%; Score 588.4; DB 6; Length 825;
Best Local Similarity 99.7%; Pred. No. 2.6e-119;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 60
|||
DB 93 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 152
|||
QY 61 TGGAGTCCAGCAGGCTTCCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 120
|||
DB 153 TGGAGTCCAGCAGGCTTCCAGAGCTACCAAGCCCTGCTCCGATCTCCCTAGACAAA 212
|||
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCCTCCGAGCATGTCTCATCATCAATAC 180
|||
DB 213 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCCTCCGAGCATGTCTCATCATCAATAC 272
|||
QY 181 ACCCTCCACAGCTGCAGCTGCATTTGGCTTGGCTCCGCCCCCTGCCCCCGAG 240
|||
DB 273 ACCCTCCACAGCTGCAGCTGCATTTGGCTTGGCTCCGCCCCCTGCCCCCGAG 332
|||
QY 241 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTTCAGCCACCATTTGGCTCTATCTCAGG 300
|||
DB 333 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTTCAGCCACCATTTGGCTCTATCTCAGG 392
|||
QY 301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTGACTCCCTTGGC 360
|||
DB 393 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAAATCCAGTGACTCCCTTGGC 452
|||
QY 361 CTCGAGATGAAGTGCACCCAGCGCTGATCCAGTCTTCTTAGAGCTCTGAGCTCCCGG 420
|||
```



```

Db      453 CTCGAGATGAAGTCCACCCAGCCCTGATCAGTCTTCTTAGAAGCTCTGAGCTCCCGG 512
Qy      421 TACTTGGGGGACTCTGGGCTGGATGACTTCTTTTGGACATTCAGACATCTGCGGTAGAA 480
Db      513 TACTTGGGGGACTCTGGGCTGGATGACTTCTTTCTGGACATTCAGACATCTGCGGTAGAA 572
Qy      481 AAGGAGCTGACGCGGCCCCCAGAGCTCTCTCAACACTTCTTGTGCCCCAGGTTCT 540
Db      573 AAGGAGCTGACGCGGCCCCCAGAGCTCTCTCAACACTTCTTGTGCCCCAGGTTCT 632
Qy      541 TGGGAGTGGAACTGAATCGATCAGATCAGATCGGAAATCATTTCTGGGGTCTCTAA 591
Db      633 TGGGAGTGGAACTGAATCGATCAGATCAGATCGGAAATCATTTCTGGGGTCTCTAA 683

RESULT 13
BD126646
LOCUS   BD126646
DEFINITION
Primer for synthesizing full-length cDNA and use thereof.
ACCESSION
BD126646
VERSION
JP 2002017375-A/2077.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 825)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
TITLE
Primer for synthesizing full-length cDNA and use thereof
JOURNAL
Patient: JP 2002017375-A 2077 22-JAN-2002;
HELIIX RESEARCH INSTITUTE
COMMENT
OS Homo sapiens (human)
PN JP 2002017375-A/2077
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI ISHII,
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI,HISASHI KOGA
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FT source
1. .825
/organism='Homo sapiens (human)'.
FT Location/Qualifiers
1. .825
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN
Query Match 99.6%; Score 588.4; DB 6; Length 825;
Best Local Similarity 99.7%; Pred No. 2.6e-119;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 60
Db      93 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTTGGAG 152
Qy      61 TGGAGTCAGACGGCTTCAGAGCTACAGCAGCCCTGCTCCGATCTCCCTAGACAAA 120
Db      153 TGGAGTCCAGAGGCGCTTCAGAGCTACAGCAGCCCTGCTCCGATCTCCCTAGACAAA 212
Qy      121 GTCCAGCCGAGCTGGGCCCCCGAGCACCCAGCTCCGAGCATGCTCTCATCATTAAC 180
Db      213 GTCCAGCGAGCTGGGCCCCCGAGCACCCAGCTCCGAGCATGCTCTCATCATTAAC 272

```

```

Qy      191 ACCCTCAACAGCTGCAGGCTGCACCTTGGCTCGGCTCCGCCCCCTGCGCTGCCCGCCCGAG 240
Db      273 ACCCTCAACAGCTGCAGGCTGCACCTTGGCTCGGCTCCGCCCCCTGCGCTGCCCGCCCGAG 332
Qy      241 CCCCTCTTCTCTGGGGAGGAGGATTTCTCCCTGTGTCAGCCACCATTTGGGTCTATCTCTCAGG 300
Db      333 CCCCTCTTCTCTGGGGAGGAGGATTTCTCCCTGTGTCAGCCACCATTTGGGTCTATCTCTCAGG 392
Qy      301 GAGCTGGACACCTCCATGATGGGACTGAGCCCTCAGAAATCCAGATCCAGTCCCTTTGGC 360
Db      393 GAGCTGGACACCTCCATGATGGGACTGAGCCCTCAGAAATCCAGATCCAGTCCCTTTGGC 452
Qy      361 CTCAGATGAAGTCCACCCAGCTGATCAGTCTTCTTAGAGCTCTGAGCTCCCGG 420
Db      453 CTCAGATGAAGTCCACCCAGCTGATCAGTCTTCTTAGAGCTCTGAGCTCCCGG 512
Qy      421 TACTTGGGGGACTCTGGGCTGGATGACTTCTTTCTGGACATTCAGACATCTGCGGTAGAA 480
Db      513 TACTTGGGGGACTCTGGGCTGGATGACTTCTTTCTGGACATTCAGACATCTGCGGTAGAA 572
Qy      481 AAGGAGCTGTCACGCGGCCCCCAGAGCTCTCTCAACACTTCTTGTGCCCCAGGTTCT 540
Db      573 AAGGAGCTGTCACGCGGCCCCCAGAGCTCTCTCAACACTTCTTGTGCCCCAGGTTCT 632
Qy      541 TGGGAGTGGAACTGAATCGATCAGATCAGATCGGAAATCATTTCTGGGGTCTCTAA 591
Db      633 TGGGAGTGGAACTGAATCGATCAGATCAGATCGGAAATCATTTCTGGGGTCTCTAA 683

RESULT 14
AC021625 164652 bp DNA linear HTG 28-MAR-2000
LOCUS   AC021625
DEFINITION
Homo sapiens clone RP11-384E6, WORKING DRAFT SEQUENCE, 34 unordered
pieces.
ACCESSION
AC021625
VERSION
AC021625.2 GI:7331453
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 164652)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-384E6
Unpublished
2 (bases 1 to 164652)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Bada,A.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneses,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
TITLE
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
On Mar 28, 2000 this sequence version replaced gi:6705474.
COMMENT
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

```

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information -----
 Center project name: L5880
 Center clone name: 384_E6
 ----- Summary Statistics -----
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 145359 bases at least Q40
 Consensus quality: 155420 bases at least Q30
 Consensus quality: 159351 bases at least Q20
 Insert size: 161352; sum-of-contigs
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 34 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1083: contig of 1083 bp in length
 * 1084 1183: gap of 100 bp
 * 1184 2479: contig of 1296 bp in length
 * 2480 2579: gap of 100 bp
 * 2580 4311: contig of 1732 bp in length
 * 4312 4411: gap of 100 bp
 * 4412 5951: contig of 1540 bp in length
 * 5952 6051: gap of 100 bp
 * 6052 7761: contig of 1710 bp in length
 * 7762 7861: gap of 100 bp
 * 7862 9630: contig of 1769 bp in length
 * 9631 9730: gap of 100 bp
 * 9731 11763: contig of 2033 bp in length
 * 11764 11863: gap of 100 bp
 * 11864 14294: contig of 2431 bp in length
 * 14295 14394: gap of 100 bp
 * 14395 16731: contig of 2337 bp in length
 * 16732 16831: gap of 100 bp
 * 16832 20280: contig of 3448 bp in length
 * 20280 20379: gap of 100 bp
 * 20380 23751: contig of 3372 bp in length
 * 23752 23851: gap of 100 bp
 * 23852 26903: contig of 3052 bp in length
 * 26904 27003: gap of 100 bp
 * 27004 31228: contig of 4225 bp in length
 * 31229 31328: gap of 100 bp
 * 31329 35336: contig of 4008 bp in length
 * 35337 35436: gap of 100 bp
 * 35437 38469: contig of 3033 bp in length
 * 38470 38569: gap of 100 bp
 * 38570 42643: contig of 4074 bp in length
 * 42644 42743: gap of 100 bp
 * 42744 47087: contig of 4344 bp in length
 * 47088 47187: gap of 100 bp
 * 47188 51475: contig of 4288 bp in length
 * 51476 51576: gap of 100 bp
 * 51576 56093: contig of 4518 bp in length
 * 56094 56193: gap of 100 bp
 * 56194 59387: contig of 3194 bp in length
 * 59388 59487: gap of 100 bp
 * 59488 63987: contig of 4500 bp in length
 * 63988 64087: gap of 100 bp
 * 64088 69347: contig of 5260 bp in length
 * 69348 74268: contig of 4821 bp in length
 * 74269 74368: gap of 100 bp
 * 74369 79365: contig of 4997 bp in length
 * 79366 79466: gap of 100 bp
 * 79466 83211: contig of 3746 bp in length
 * 83212 83311: gap of 100 bp
 * 83312

* 83312 88101: contig of 4790 bp in length
 * 88102 88201: gap of 100 bp
 * 88202 93785: contig of 5584 bp in length
 * 93786 93885: gap of 100 bp
 * 93886 100407: contig of 6522 bp in length
 * 100408 100507: gap of 100 bp
 * 100508 108158: contig of 7651 bp in length
 * 108159 108258: gap of 100 bp
 * 108259 115631: contig of 7373 bp in length
 * 115632 115731: gap of 100 bp
 * 115732 125488: contig of 9757 bp in length
 * 125489 136013: contig of 10425 bp in length
 * 125589 136113: gap of 100 bp
 * 136014 149110: contig of 12997 bp in length
 * 136114 149210: gap of 100 bp
 * 149111 149211: contig of 15442 bp in length.
 * 149211 164652: contig of 15442 bp in length.
 *
 FEATURES
 source
 1. 164652
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-384E6"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1083
 /note="assembly_fragment"
 1184. 2479
 /note="assembly_fragment"
 2580. 4311
 /note="assembly_fragment"
 4412. 5951
 /note="assembly_fragment"
 6052. 7761
 /note="assembly_fragment"
 7862. 9630
 /note="assembly_fragment"
 9731. 11763
 /note="assembly_fragment"
 11864. 14294
 /note="assembly_fragment"
 14395. 16731
 /note="assembly_fragment"
 16832. 20279
 /note="assembly_fragment"
 20380. 23751
 /note="assembly_fragment"
 23852. 26903
 /note="assembly_fragment"
 27004. 31228
 /note="assembly_fragment"
 31329. 35336
 /note="assembly_fragment"
 35437. 38469
 /note="assembly_fragment"
 38570. 42643
 /note="assembly_fragment"
 42744. 47087
 /note="assembly_fragment"
 47188. 51475
 /note="assembly_fragment"
 51576. 56093
 /note="assembly_fragment"
 56194. 59387
 /note="assembly_fragment"
 59488. 63987
 /note="assembly_fragment"
 64088. 69347
 /note="assembly_fragment"
 69448. 74268
 /note="assembly_fragment"
 74369. 79365
 /note="assembly_fragment"
 79466. 83211
 /note="assembly_fragment"
 misc_feature
 1. 1083
 /note="assembly_fragment"
 misc_feature
 1184. 2479
 /note="assembly_fragment"
 misc_feature
 2580. 4311
 /note="assembly_fragment"
 misc_feature
 4412. 5951
 /note="assembly_fragment"
 misc_feature
 6052. 7761
 /note="assembly_fragment"
 misc_feature
 7862. 9630
 /note="assembly_fragment"
 misc_feature
 9731. 11763
 /note="assembly_fragment"
 misc_feature
 11864. 14294
 /note="assembly_fragment"
 misc_feature
 14395. 16731
 /note="assembly_fragment"
 misc_feature
 16832. 20279
 /note="assembly_fragment"
 misc_feature
 20380. 23751
 /note="assembly_fragment"
 misc_feature
 23852. 26903
 /note="assembly_fragment"
 misc_feature
 27004. 31228
 /note="assembly_fragment"
 misc_feature
 31329. 35336
 /note="assembly_fragment"
 misc_feature
 35437. 38469
 /note="assembly_fragment"
 misc_feature
 38570. 42643
 /note="assembly_fragment"
 misc_feature
 42744. 47087
 /note="assembly_fragment"
 misc_feature
 47188. 51475
 /note="assembly_fragment"
 misc_feature
 51576. 56093
 /note="assembly_fragment"
 misc_feature
 56194. 59387
 /note="assembly_fragment"
 misc_feature
 59488. 63987
 /note="assembly_fragment"
 misc_feature
 64088. 69347
 /note="assembly_fragment"
 misc_feature
 69448. 74268
 /note="assembly_fragment"
 misc_feature
 74369. 79365
 /note="assembly_fragment"
 misc_feature
 79466. 83211
 /note="assembly_fragment"

```
misc_feature /note="assembly_fragment"  
8312..88101  
misc_feature /note="assembly_fragment"  
88202..93785  
misc_feature /note="assembly_fragment"  
93886..100407  
misc_feature /note="assembly_fragment"  
clone_end:17  
vector_side:right  
misc_feature 100508..108158  
misc_feature /note="assembly_fragment"  
108259..115631  
  
Query Match 98.9%; Score 584.6; DB 2; Length 164652;  
Best Local Similarity 99.3%; Pred. No. 9e-119;  
Matches 587; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 ATGGAGGAGGCTTGAAGAGGAACACTCTGATTTGGAGAGGAGGAGAGGTGGGAG 60  
DB 113509 ATGGTGGGAGGCTTGAAGAGGAACACTCTGATTTGGAGAGGAGGAGAGGTGGGAG 113568  
  
QY 61 TGGAGTCCAGCAGGCGCTTCAGAGCTACCAAGAGCCCTGCTCGCATCTCCCTAGACAAA 120  
DB 113569 TGGAGTCCAGCAGGCGCTTCAGAGCTACCAAGAGCCCTGCTCGCATCTCCCTAGACAAA 113628  
  
QY 121 GTCCAGCGCAGCGCTGGGCCCCCGAGCACCCAGCTCCCGCAGCATGTCTCATCATTAAC 180  
DB 113629 GTCCAGCGCAGCGCTGGGCCCCCGAGCACCCAGCTCCCGCAGCATGTCTCATCATTAAC 113688  
  
QY 181 ACCCTCCACAGCTCGAGGCTGCATCTGCGCTGGCTCCCGCCCTGCGCCCGCCCGAG 240  
DB 113689 ACCCTCCACAGCTCGAGGCTGCATCTGCGCTGGCTCCCGCCCTGCGCCCGCCCGAG 113748  
  
QY 241 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCCACCATGGCTCTATCTCAGG 300  
DB 113749 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCCACCATGGCTCTATCTCAGG 113808  
  
QY 301 GAGCTGGACACCTCCATGATGGGACTGAGCCCGCTCAGAAATCCAGTGCATCCCTTTGGC 360  
DB 113809 GAGCTGGACACCTCCATGATGGGACTGAGCCCGCTCAGAAATCCAGTGCATCCCTTTGGC 113868  
  
QY 361 CTCAGAGTGAAGTCCACCCAGCTGATGAGTCTTCTTGGACATTTGACATCTCGGTTAGAA 420  
DB 113869 CTCAGAGTGAAGTCCACCCAGCTGATGAGTCTTCTTGGACATTTGACATCTCGGTTAGAA 113928  
  
QY 421 TACTTGGGGGACTCTGGGCTGGATGACTTCTTCTTGGACATTTGACATCTCGGTTAGAA 480  
DB 113929 TACTTGGGGGACTCTGGGCTGGATGACTTCTTCTTGGACATTTGACATCTCGGTTAGAA 113988  
  
QY 481 AAGGAGCTTCACGGGCCCCCAGAGCCTCTCAACCTCTTCTGTGCCCCAGGTTCT 540  
DB 113989 AAGGAGCTTCACGGGCCCCCAGAGCCTCTCAACCTCTTCTGTGCCCCAGGTTCT 114048  
  
QY 541 TGGGAGTGAAGTGAATGATGATCATCATCTGGAATCATCTCGGGTCTCTAA 591  
DB 114049 TGGGAGTGAAGTGAATGATGATCATCATCTGGAATCATCTCGGGTCTCTAA 114099
```

```
RESULT 15  
LOCUS CQ463286  
DEFINITION Sequence 6064 from Patent WO0192581.  
ACCESSION CQ463286  
VERSION CQ463286.1 GI:41428905  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Algate,P.A., Harlocker,S.L. and Jones,R.  
TITLE Compositions and methods for the therapy and diagnosis of  
ovariancancer
```

```
JOURNAL Patent: WO 0192581-A 6064 06-DEC-2001;  
CORIXA CORPORATION (US)  
FEATURES Location/Qualifiers  
source 1..506  
/organism="Homo sapiens"  
/mol_type="unassigned DNA"  
/db_xref="taxon:9606"  
  
ORIGIN  
  
Query Match 80.0%; Score 473; DB 6; Length 506;  
Best Local Similarity 98.6%; Pred. No. 7.4e-94;  
Matches 488; Conservative 0; Mismatches 5; Indels 2; Gaps 1;  
  
QY 60 GTGGAGTCCAGCAGCGCTTCAGAGCTACCAAGAGCCCTGTCCGATCTCCCTAGACAA 119  
DB 14 GCGGAGTCCAGCAGCGCTTCAGAGCTACCAAGAGCCCTGTCCGATCTCCCTAGACAA 73  
  
QY 120 AGTCCAGCGCAGCGCTGGGCCCCCGAGCACCCAGCTCCCGAGCATGTCTCATCATTA 179  
DB 74 AGTCCAGCGC--CCTGGGCCCCCGAGCACCCAGCTCCCGAGCATGTCTCATCATTA 131  
  
QY 180 CACCTCTCAACAGCTGCAAGGCTGCATCTCGCTGGCTCCCGCCCTGCGCCCGCCGA 239  
DB 132 CACCTCTCAACAGCTGCAAGGCTGCATCTCGCTGGCTCCCGCCCTGCGCCCGCCGA 191  
  
QY 240 GCCCTCTTCTGGGCGAGGAGGATTTCTCTGTGAGCCACCATTTGGCTCTATCTCAG 299  
DB 192 GCCCTCTTCTGGGCGAGGAGGATTTCTCTGTGAGCCACCATTTGGCTCTATCTCAG 251  
  
QY 300 GGAGCTGGACACCTCCATGATGGGACTGAGCCCGCTCAGATCCAGTGCATCCCTTGG 359  
DB 252 GGAGCTGGACACCTCCATGATGGGACTGAGCCCGCTCAGATCCAGTGCATCCCTTGG 311  
  
QY 360 CCTCCAGAAATGAAGTGCACCCAGCTGATCCAGTCTTTTAGAAGCTCTGAGCTCCCG 419  
DB 312 CCTCCAGAAATGAAGTGCACCCAGCTGATCCAGTCTTTTAGAAGCTCTGAGCTCCCG 371  
  
QY 420 GTACTTGGGGGACTCTGGCTGGATGACTTTCTTTTGGACATTTGACATCTGCGGTAGA 479  
DB 372 GTACTTGGGGGACTCTGGCTGGATGACTTTCTTTTGGACATTTGACATCTGCGGTAGA 431  
  
QY 480 AAAGGAGCTTCACGGGCCCCCAGAGCCTCTCAACCTCTTCTGTGCCCCAGGTTTC 539  
DB 432 AAAGGAGCTTCACGGGCCCCCAGAGCCTCTCAACCTCTTCTGTGCCCCAGGTTTC 491  
  
QY 540 TTGGGAGTGGAAATGA 554  
DB 492 TTGGGAGTGGAAATGA 506
```

Search completed: April 26, 2005, 09:13:33
Job time : 2755 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2005, 03:13:57 ; Search time 425 Seconds
(without alignments)
8231.923 Million cell updates/sec

Title: US-10-069-386A-1
Perfect score: 591
Sequence: 1 atggaggaggcttgaaagag.....aaatcattctggggctcctaa 591

Scoring table: IDENTITY_NUC

Gapop 10.0 , Capext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_16Dec04.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	591	100.0	591	4	Aaf28052
2	589.4	99.7	1325	4	Aaf28052 Replicati
3	589.4	99.7	1325	4	Aak94884 Human ful
4	589.4	99.7	1325	12	Adl32051 Full leng
5	589.4	99.7	1515	13	Adg85787 Human tum
6	588.4	99.6	825	4	Adg33127 Human DNA
7	588.4	99.6	825	4	Aak91661 Human cDN
8	588.4	99.6	825	4	Aak93617 Human cDN
9	588.4	99.6	825	12	Adl30844 3' end of
10	578.4	97.9	598	12	Adl28088 5' end of
11	576.2	97.5	1596	6	Aah34147 Human gen
12	473	80.0	506	6	Abl83086 Human ova
13	186.6	31.6	205	6	Abl3261 Human ova
14	166	28.1	526	12	Ach77983 Human gen
15	47.4	8.0	2082	6	Aal41497 Mouse hae
16	46.4	7.9	51552	6	Aas96607 DNA encod
17	42.6	7.2	1603	8	Acc51095 Human Sho
18	42.4	7.2	511	12	Ach91786 Human gen
19	42.4	7.2	600	6	Abq52497 Oligonuc
20	42.4	7.2	600	6	Abq52496 Oligonuc

C	21	42.4	7.2	1850	8	ADA70325	Ada70325 Rice gene
C	22	41.8	7.1	32329	12	ADO51695	ADO51695 Streptomy
C	23	41.4	7.0	825	12	ADO00165	ADO00165 Novel hum
C	24	41.4	7.0	825	12	ADN98596	ADN98596 Novel hum
C	25	41	6.9	1925	2	AAX90924	Aax90924 Epstein B
C	26	41	6.9	1926	3	AAA50254	Aas50254 Epstein B
C	27	41	6.9	1926	4	Aaf82902	Aaf82902 ERV tethe
C	28	41	6.9	1926	10	ADK65580	Adk65580 Human her
C	29	41	6.9	2580	3	AAA75454	Aaa75454 Nucleotid
C	30	41	6.9	2580	6	AAI64275	Aai64275 Epstein-B
C	31	41	6.9	5452	2	AAX90923	Aax90923 Anti-sens
C	32	41	6.9	8705	2	AZ233778	Aaz233778 Vector ps
C	33	41	6.9	8705	12	ADM10659	Adm10659 Expressio
C	34	41	6.9	9482	12	ADP64415	Adp64415 Vector pc
C	35	41	6.9	9600	2	AAV21683	Aav21683 Vector pl
C	36	41	6.9	10285	6	ABS71027	Abst71027 pCEP-Xa-F
C	37	41	6.9	10285	6	ABS66453	Abst66453 Plasmid p
C	38	41	6.9	10330	12	ADL67154	Adl67154 Plasmid p
C	39	41	6.9	10380	2	AZ22248	Aaz22248 Nucleotid
C	40	41	6.9	10477	12	ADL67152	Adl67152 Plasmid p
C	41	41	6.9	10516	12	ADL67150	Adl67150 Plasmid p
C	42	41	6.9	10561	12	ADL67148	Adl67148 Plasmid p
C	43	41	6.9	10596	2	AAQ51731	Aaq51731 Plasmid p
C	44	41	6.9	10596	2	AAX15650	Aax15650 Nucleotid
C	45	41	6.9	10596	2	AAT40348	Aat40348 Plasmid p

ALIGNMENTS

RESULT 1
AAF28052
ID AAF28052 standard; DNA; 591 BP.
XX
AC AAF28052;
XX
DT 23-MAY-2001 (first entry)
XX
DE Replication protein A binding transcriptional activator 1 RBT1 gene.
XX
KW RBT1; replication protein A binding transcriptional activator 1; RPA32;
KW gene therapy; apoptosis; cancer; leukaemia; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..591
FT /tag= a
FT /product= "RBT1"
XX
PN WO200114546-A2.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-CA000948.
XX
PR 19-AUG-1999; 99US-0149472P.
XX
PA (TRAN-) CENT TRANSLATIONAL RES IN CANCER.
XX
PI Alaoui-Jamali MA, Cho JM;
XX
DR WPI; 2001-218447/22.
XX
DR P-PSDB; AAB35402.
XX
PT Novel replication protein A binding transcriptional activator 1 gene,
PT useful for treating neoplastic disorders such as cancer and in gene
PT therapy.
XX
PS Claim 1; Fig 1; 16pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC replication protein A binding transcriptional activator 1 (RBT1). The

CC protein is capable of inducing apoptosis. The sequences are useful in the
CC gene therapy and other methods of treatment of cancer, including
CC leukaemias. The present sequence is the RBT1 coding sequence
XX

SQ Sequence 591 BP; 123 A; 194 C; 156 G; 118 T; 0 U; 0 Other;
Query Match 100.0%; Score 591; DB 4; Length 591;
Best Local Similarity 100.0%; Pred. No. 4.3e-135;
Matches 591; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTGTGATTTGGGAAGAGGAGGAGGTGGGAG 60
DB |||||
QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTGTGATTTGGGAAGAGGAGGAGGTGGGAG 60
DB |||||
QY 61 TGGAGTCAGAGGCGCTTCAGAGCTACAGCAAGCGCTGCTCGCATCTCCCTAGACAA 120
DB |||||
QY 61 TGGAGTCAGAGGCGCTTCAGAGCTACAGCAAGCGCTGCTCGCATCTCCCTAGACAA 120
DB |||||
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCTCCGAGGCATGTCTCATCATTAAC 180
DB |||||
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCTCCGAGGCATGTCTCATCATTAAC 180
DB |||||
QY 181 ACCCTCCAAACAGCTGAGGCTGACCTTGCCTGGCTCCGCCCCCTGCCCCCGAG 240
DB |||||
QY 181 ACCCTCCAAACAGCTGAGGCTGACCTTGCCTGGCTCCGCCCCCTGCCCCCGAG 240
DB |||||
QY 241 CCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
DB |||||
QY 241 CCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
DB |||||
QY 301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTGAATCTTGGC 360
DB |||||
QY 301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTGAATCTTGGC 360
DB |||||
QY 361 CTCAGAGATGAAGTGCACCCAGCCTGATCAGTCTTTCTTGAAGAGCTGTGAGTCCCGG 420
DB |||||
QY 361 CTCAGAGATGAAGTGCACCCAGCCTGATCAGTCTTTCTTGAAGAGCTGTGAGTCCCGG 420
DB |||||
QY 421 TACTTTGGGGAGCTCTGGGCTGGATGACTTCTTTCTGGACATTTGACACATCTCGGTA 480
DB |||||
QY 421 TACTTTGGGGAGCTCTGGGCTGGATGACTTCTTTCTTGGACATTTGACACATCTCGGTA 480
DB |||||
QY 481 AAGGAGCTTGCAGCGGCCCCCAGAGCCTCCTCACAACCTCTTGTGCCCCCAGGTTCT 540
DB |||||
QY 481 AAGGAGCTTGCAGCGGCCCCCAGAGCCTCCTCACAACCTCTTGTGCCCCCAGGTTCT 540
DB |||||
QY 541 TGGGAGTGAATGAATGAATGGATCACATCATATGGAATCATTTCTGGGTCTCTAA 591
DB |||||
QY 541 TGGGAGTGAATGAATGAATGGATCACATCATATGGAATCATTTCTGGGTCTCTAA 591
DB |||||

RESULT 2
AAK94884
ID AAK94884 standard; cDNA; 1325 BP.

XX AAK94884;
AC AAK94884;
XX

DT 06-NOV-2001 (first entry)

DE Human full-length cDNA, SEQ ID NO: 4084.

KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

OS Homo sapiens.

XX EP1130094-A2.

PD 05-SEP-2001.

XX 07-JUL-2000; 2000EP-00114089.

XX 08-JUL-1999; 99JP-0019486.

PR 11-JAN-2000; 2000JP-00118774.

PR 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.
DR P-PSDB; AM93922.

XX 830 Primers useful for synthesizing full length cDNA clones and their use
in genetic manipulation.

PS Claim 8; SEQ ID NO 4084; 1380pp + Sequence Listing; English.

CC The invention relates to primers for synthesising full length cDNA
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesising the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO

XX Sequence 1325 BP; 263 A; 371 C; 359 G; 332 T; 0 U; 0 Other;

SQ Query Match 99.7%; Score 589.4; DB 4; Length 1325;
Best Local Similarity 99.8%; Pred. No. 1.3e-134;
Matches 590; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGAGGAGGCTTGAAGAGGAAACACTGTGATTTGGGAAGAGGAGGAGGTGGGAG 60
DB |||||
QY 93 ATGGTGGGAGGCTTGAAGAGGAAACACTGTGATTTGGGAAGAGGAGGAGGTGGGAG 152
DB |||||
QY 61 TGGAGTCAGAGGCGCTTCAGAGCTACAGCAAGCGCTGCTCGCATCTCCCTAGACAA 120
DB |||||
QY 153 TGGAGTCAGAGGCGCTTCAGAGCTACAGCAAGCGCTGCTCGCATCTCCCTAGACAA 212
DB |||||
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCTCCGAGGCATGTCTCATCATTAAC 180
DB |||||
QY 213 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCTCCGAGGCATGTCTCATCATTAAC 272
DB |||||
QY 181 ACCCTCCAAACAGCTGAGGCTGACCTTGCCTGGCTCCGCCCCCTGCCCCCGAG 240
DB |||||
QY 273 ACCCTCCAAACAGCTGAGGCTGACCTTGCCTGGCTCCGCCCCCTGCCCCCGAG 332
DB |||||
QY 241 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
DB |||||
QY 333 CCCCTCTTCTGGGCGAGGAGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 392
DB |||||
QY 301 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTGAATCTTGGC 360
DB |||||
QY 393 GAGCTGGACACCTCCATGGATGGGACTGAGCCCCCTCAGAATCCAGTGAATCTTGGC 452
DB |||||
QY 361 CTCAGAGATGAAGTGCACCCAGCCTGATCAGTCTTTCTTGAAGAGCTGTGAGTCCCGG 420
DB |||||
QY 453 CTCAGAGATGAAGTGCACCCAGCCTGATCAGTCTTTCTTGAAGAGCTGTGAGTCCCGG 512
DB |||||
QY 421 TACTTTGGGGAGCTCTGGGCTGGATGACTTCTTTCTTGGACATTTGACACATCTCGGTA 480
DB |||||
QY 513 TACTTTGGGGAGCTCTGGGCTGGATGACTTCTTTCTTGGACATTTGACACATCTCGGTA 572
DB |||||
QY 481 AAGGAGCTTGCAGCGGCCCCCAGAGCCTCCTCACAACCTCTTGTGCCCCCAGGTTCT 540
DB |||||
QY 573 AAGGAGCTTGCAGCGGCCCCCAGAGCCTCCTCACAACCTCTTGTGCCCCCAGGTTCT 632
DB |||||
QY 541 TGGGAGTGAATGAATGAATGGATCACATCATATGGAATCATTTCTGGGTCTCTAA 591
DB |||||
QY 633 TGGGAGTGAATGAATGAATGGATCACATCATATGGAATCATTTCTGGGTCTCTAA 683
DB |||||

Db 412 GTCCAGCGCAGCTGGGCCCCCGAGCACCAGCCTCCCGAGGCATGTCTCATCATAC 471
Qy 181 ACCCTCCAAAGCTGAGCTGACCTTCCCTGGCTCCCGCCCTCCCTGCCCCCGAG 240
Db 472 ACCCTCCAAAGCTGAGCTGACCTTCCCTGGCTCCCGCCCTCCCTGCCCCCGAG 531
Qy 241 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCACCATTGGCTCTATCTCAGG 300
Db 532 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCACCATTGGCTCTATCTCAGG 591
Qy 301 GAGCTGGACACCTCCATGATGGGAGTCCAGCCCTCCAGAACTCCAGTCACTCCCTTGGC 360
Db 592 GAGCTGGACACCTCCATGATGGGAGTCCAGCCCTCCAGAACTCCAGTCACTCCCTTGGC 651
Qy 361 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTTTGAAGAGCTGTAGTCCCGG 420
Db 652 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTTTGAAGAGCTGTAGTCCCGG 711
Qy 421 TACTTGGGGAGCTCTGGGCTGGATGACTTCTTCTGGACATTTGACATCTCGGCTAGAA 480
Db 712 TACTTGGGGAGCTCTGGGCTGGATGACTTCTTCTGGACATTTGACATCTCGGCTAGAA 771
Qy 481 AAGGAGCTTGCACGGGCCCCCAGAGCCTCTCTCAACCTCTTCTGTGCCCTCAGTTCT 540
Db 772 AAGGAGCTTGCACGGGCCCCCAGAGCCTCTCTCAACCTCTTCTGTGCCCTCAGTTCT 831
Qy 541 TGGGAGTGAATGAATGGATCAGATCAGATCATATGGAATCATTTCTGGGTCTCTAA 591
Db 832 TGGGAGTGAATGAATGGATCAGATCAGATCATATGGAATCATTTCTGGGTCTCTAA 882

RESULT 6
AAK91661
ID AAK91661 standard; cDNA; 825 BP.
XX
AC AAK91661;
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 121.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Claim 2; SEQ ID NO 121; 1380pp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA

CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is the nucleotide sequence of the 5'-end of
CC a cDNA provided in the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in CD-
CC ROM format directly from EPO
XX
SQ Sequence 825 BP; 160 A; 247 C; 230 G; 183 T; 0 U; 5 Other;
Query Match 99.6%; Score 588.4; DB 4; Length 825;
Best Local Similarity 99.7%; Pred. No. 2e-134;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGGAGGAGGCTTGAAGAGGAAACACTCTGATTTTGGAGAGGAGGAGGTTGGGAG 60
Db 93 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTTGGAGAGGAGGAGGTTGGGAG 152
Qy 61 TGGAGTCCAGCAGGCTTTCAGAGCTTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 120
Db 153 TGGAGTCCAGCAGGCTTTCAGAGCTTACCAGCAAGCCCTGCTCCGCATCTCCCTAGACAAA 212
Qy 121 GTCCAGCGCAGCTTGGGCCCCCGAGCACCAGCCTCCCGAGGCATGTCTCATCATAC 180
Db 213 GTCCAGCGCAGCTTGGGCCCCCGAGCACCAGCCTCCCGAGGCATGTCTCATCATAC 272
Qy 181 ACCCTTCAACAGCTGCAGGCTGCACTTCGCTGGCTCCCGCCCTGCGCTGCCCCCGAG 240
Db 273 ACCCTTCAACAGCTGCAGGCTGCACTTCGCTGGCTCCCGCCCTGCGCTGCCCCCGAG 332
Qy 241 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCACCATTGGCTCTATCTCAGG 300
Db 333 CCCCTCTTCTGGGCGAGGAGGATTTCTCCCTGTGAGCACCATTGGCTCTATCTCAGG 392
Qy 301 GAGCTGGACACCTCCATGGATGGGAGTCCAGCCCTCCAGAACTCCAGTCACTCCCTTGGC 360
Db 393 GAGCTGGACACCTCCATGGATGGGAGTCCAGCCCTCCAGAACTCCAGTCACTCCCTTGGC 452
Qy 361 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTTTGAAGAGCTGTAGTCCCGG 420
Db 453 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTTTGAAGAGCTGTAGTCCCGG 512
Qy 421 TACTTGGGGAGCTCTGGGCTGGATGACTTCTTCTGGACATTTGACATCTCGGCTAGAA 480
Db 513 TACTTGGGGAGCTCTGGGCTGGATGACTTCTTCTGGACATTTGACATCTCGGCTAGAA 572
Qy 481 AAGGAGCTTGCACGGGCCCCCAGAGCCTCTCTCAACCTCTTCTGTGCCCTCAGTTCT 540
Db 573 AAGGAGCTTGCACGGGCCCCCAGAGCCTTCTCAACCTCTTCTGTGCCCTCAGTTCT 632
Qy 541 TGGGAGTGAATGAATGGATCAGATCAGATCATATGGAATCATTTCTGGGTCTCTAA 591
Db 633 TGGGAGTGAATGAATGGATCAGATCAGATCATATGGAATCATTTCTGGGTCTCTAA 683

RESULT 7
AAK93617
ID AAK93617 standard; cDNA; 825 BP.
XX
AC AAK93617;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA clone representative sequence, SEQ ID NO: 2077.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.

XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
PT
XX Example 11; SEQ ID NO 2077; 1380pp + Sequence Listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence was used as the representative sequence
CC from a human clone which was used in homology searches to identify the
CC clone. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in CD-ROM format directly from
CC EPO
XX
XX Sequence 825 BP; 160 A; 247 C; 230 G; 183 T; 0 U; 5 Other;
PA
Query Match 99.6%; Score 588.4; DB 4; Length 825;
Best Local Similarity 99.7%; Pred. No. 2e-134;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 60
DB |||||
93 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 152
XX
QY 61 TGGAGTCCAGCAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 120
DB |||||
153 TGGAGTCCAGCAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 212
XX
QY 121 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCTCCGAGCATGTCTCATCCATAAC 180
DB |||||
213 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCTCCGAGCATGTCTCATCCATAAC 272
XX
QY 181 ACCCTCCAGCAGCTGGGCCCCGAGCACCAGCCTCCGAGCATGTCTCATCCATAAC 240
DB |||||
273 ACCCTCCAGCAGCTGGGCCCCGAGCACCAGCCTCCGAGCATGTCTCATCCATAAC 332
XX
QY 241 CCCCTCTTCTGGGGAGGAGGATTTCTCCCTGTGTCAGCCACCATTTGGCTCTATCTCAGG 300
DB |||||
333 CCCCTCTTCTGGGGAGGAGGATTTCTCCCTGTGTCAGCCACCATTTGGCTCTATCTCAGG 392
XX
QY 301 GAGCTGGACACCTCATGATGGGACTGAGCCCCCTCAGAAATCCAGTCACTCCCTTGGC 360
DB |||||
393 GAGCTGGACACCTCATGATGGGACTGAGCCCCCTCAGAAATCCAGTCACTCCCTTGGC 452
XX
QY 361 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTCTTGAAGAGCTCTGAGCTCCCGG 420
DB |||||
453 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTCTTGAAGAGCTCTGAGCTCCCGG 512
XX
QY 421 TACTTTGGGGAGCTCTGGCTGGATGACTTTCTTTCTGGACATTTGACACATCTGCGGTAGAA 480
DB |||||
513 TACTTTGGGGAGCTCTGGCTGGATGACTTTCTTTCTGGACATTTGACACATCTGCGGTAGAA 572
XX
QY 481 AAGGAGCCTGACAGGGCCCCCAGAGCCTCTCTCAACACCTTTCTGTGCCCCCAGGTTCT 540
DB |||||
573 AAGGAGCCTGACAGGGCCCCCAGAGCCTCTCTCAACACCTTTCTGTGCCCCCAGGTTCT 632

QY 541 TGGGAGTGAATGAACCTGGATCAGATCATGGAATCATTTCTGGGGTCTTAA 591
DB |||||
633 TGGGAGTGAATGAACCTGGATCAGATCATGGAATCATTTCTGGGGTCTTAA 683
XX
RESULT 8
ADL30044
ID ADL30044 standard; cDNA; 825 BP.
XX
XX ADL30044;
AC
XX 20-MAY-2004 (first entry)
DT
XX 3' end of a representative human cDNA cluster SeqID 2077.
DE
XX human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method; ss.
KW
XX Homo sapiens.
OS
XX EPI396543-A2.
PN
XX 10-MAR-2004.
PD
XX 07-JUL-2000; 2003EP-00025638.
PF
XX 08-JUL-1999; 95JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2004-204755/20.
XX
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
PT
XX Example 18; SEQ ID NO 2077; 1340pp; English.
PS
XX This invention relates to a novel primers useful for synthesising full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polynucleotide sequence is the
CC 3' end of a representative human DNA cluster of the invention.
XX
XX Sequence 825 BP; 160 A; 247 C; 230 G; 183 T; 0 U; 5 Other;
SQ
Query Match 99.6%; Score 588.4; DB 12; Length 825;
Best Local Similarity 99.7%; Pred. No. 2e-134;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 ATGGAGGGAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 60
DB |||||
93 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 152
XX
QY 61 TGGAGTCCAGCAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 120
DB |||||
153 TGGAGTCCAGCAGGCTTGAAGAGGAAACACTCTGATTTGGGAAGAGGAGGAGGTGGGAG 212
XX
QY 121 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCTCCGAGCATGTCTCATCCATAAC 180
DB |||||
213 GTCCAGCGCAGCTGGGCCCCGAGCACCAGCCTCCGAGCATGTCTCATCCATAAC 272
XX
QY 181 ACCCTCCAGCAGCTGGGCCCCGAGCACCAGCCTCCGAGCATGTCTCATCCATAAC 240
DB |||||

Db 273 ACCCTCCACAGCTGCAGCTGCACTTCGCTGGCTCCGCCCCCTGCGCTGCCCTCCGAG 332
QY 241 CCCCTCTTCCTGGGAGGAGGATTTCTCCCTGTGTCAGCCACCAATGGCTCTATCTCAGG 300
Db 333 CCCCTCTTCCTGGGAGGAGGATTTCTCCCTGTGTCAGCCACCAATGGCTCTATCTCAGG 392
QY 301 GAGCTGGACACCTCCATGATGGGATGGGATGAGCCCTCAGAAATCCAGTCACTCCCTTGGC 360
Db 393 GAGCTGGACACCTCCATGATGGGATGGGATGAGCCCTCAGAAATCCAGTCACTCCCTTGGC 452
QY 361 CTCAGAAATGAAGTGCACCCAGCTGATCAGTCTCTTCTAGAAAGCTCTGAGCTCCCGG 420
Db 453 CTCAGAAATGAAGTGCACCCAGCTGATCAGTCTCTTCTAGAAAGCTCTGAGCTCCCGG 512
QY 421 TACTTGGGGAGCTCTGGCTGGATGACCTTTCTTCTGGAACATTCGACATCTGCGGTAGAA 480
Db 513 TACTTGGGGAGCTCTGGCTGGATGACCTTTCTTCTGGAACATTCGACATCTGCGGTAGAA 572
QY 481 AAGGAGCTTGACGGGCCCCCAGAGCTCTCTCAACCTTCTTCTGTCGCCAGGTTCT 540
Db 573 AAGGAGCTTGACGGGCCCCCAGAGCTCTCTCAACCTTCTTCTGTCGCCAGGTTCT 632
QY 541 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
Db 633 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 683

RESULT 9

ID ADL28088 standard; cDNA; 825 BP.
XX
AC ADL28088;
XX
DT 20-MAY-2004 (first entry)
XX
DE 5' end of a human cDNA molecule SeqID 121.
XX
KW human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method; ss.
OS Homo sapiens.
XX
FN EPI396543-A2.
XX
PD 10-MAR-2004.
XX
PF 07-JUL-2000; 2003EP-00025638.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
PI WPI; 2004-204755/20.
XX
PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX
PS Disclosure; SEQ ID NO 121; 1340pp; English.

XX This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3'
CC ends using the oligo-capping method. This polynucleotide sequence is the

CC 5' end of a full length human cDNA sequence of the invention.

XX
SQ Sequence 825 BP; 160 A; 247 C; 230 G; 183 T; 0 U; 5 Other;
Query Match 99.6%; Score 588.4; DB 12; Length 825;
Best Local Similarity 99.7%; Pred. No. 2e-134;
Matches 589; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGAGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGTGGGAG 60
Db 93 ATGGTGGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGTGGGAG 152
QY 61 TGGAGTCCAGCAGGCTTTCAGAGCTACAGCAGCCCTGCTCCGATCTCCCTAGACAA 120
Db 153 TGGAGTCCAGCAGGCTTTCAGAGCTACAGCAGCCCTGCTCCGATCTCCCTAGACAA 212
QY 121 GTCCAGCCAGCCTGGGCCCCCGAGCACCAGCCTCCGAGCAGCATGCTCATCATCAAC 180
Db 213 GTCCAGCCAGCCTGGGCCCCCGAGCACCAGCCTCCGAGCAGCATGCTCATCATCAAC 272
QY 181 ACCCTCCAAACAGCTGAGGCTGCATCTTCCCTGGCTCCGCCCCCTGCCCCCGAG 240
Db 273 ACCCTCCAAACAGCTGAGGCTGCATCTTCCCTGGCTCCGCCCCCTGCCCCCGAG 332
QY 241 CCCCTCTTCTGGGAGGAGGATTTCTCCCTGTGTCAGCACCATTTGGCTCTATCTCAGG 300
Db 333 CCCCTCTTCTGGGAGGAGGATTTCTCCCTGTGTCAGCACCATTTGGCTCTATCTCAGG 392
QY 301 GAGCTGGACACCTCCATGATGGGACTGAGCCCCCTCAGAAATCCAGTCACTCCCTTGGC 360
Db 393 GAGCTGGACACCTCCATGATGGGACTGAGCCCCCTCAGAAATCCAGTCACTCCCTTGGC 452
QY 361 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTTCTTAGAAGCTCTGAGCTCCCGG 420
Db 453 CTCAGAAATGAAGTGCACCCAGCCTGATCAGTCTTTCTTAGAAGCTCTGAGCTCCCGG 512
QY 421 TACTTGGGGAGCTCTGGCTGGATGACTTCTTCTGGAACATTCGACATCTGCGGTAGAA 480
Db 513 TACTTGGGGAGCTCTGGCTGGATGACTTCTTCTGGAACATTCGACATCTGCGGTAGAA 572
QY 481 AAGGAGCTTGACGGGCCCCCAGAGCTCTCTCAACCTTCTTCTGTCGCCAGGTTCT 540
Db 573 AAGGAGCTTGACGGGCCCCCAGAGCTCTCTCAACCTTCTTCTGTCGCCAGGTTCT 632
QY 541 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
Db 633 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 683

RESULT 10

ACH91703
ID ACH91703 standard; DNA; 598 BP.
XX
AC ACH91703;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human genome derived single exon probe #24898.
XX
KW Human; probe; ss; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX
OS Homo sapiens.
XX
PN US2003194704-A1.
XX
PD 16-OCT-2003.
XX
PF 03-APR-2002; 2002US-00029386.
XX
PR 03-APR-2002; 2002US-00029386.
XX
PA (PENN/) PENN S G.

(RANK/) RANK D R.
(HANZ/) HANZEL D K.
Penn SG, Rank DR, Hanzel DK;
WPI; 2004-119264/12.
New human genome-derived single exon nucleic acid probes useful for human gene expression analysis, for identifying or characterizing alternative splicing events, for assessing genomic alterations or as tools for surveying tissues.
Claim 1; SEQ ID NO 24898; 80pp; English.
The invention relates to a nucleic acid probe for measuring human gene expression, comprising any of the 27,400 fully defined nucleotide sequences in the specification, or their complements or fragments, and encoding at least 8 amino acids of any of the 6888 amino acid sequences fully defined in the specification. The probe is a single exon probe that hybridizes under high stringency conditions to a nucleic acid molecule expressed in human cells or tissues. Also included are a spatially-addressable set of single exon nucleic acid probes for measuring human gene expression (comprising a plurality of single exon nucleic acid probes cited above, where each of the plurality of probes is separately and addressably isolatable or amplifiable from the plurality), a single exon microarray for measuring human gene expression, a method of measuring human gene expression, a vector comprising the single exon probe cited above, an ORF-encoded peptide comprising at least 8 contiguous amino acids of any of the above-mentioned amino acid sequences (optionally with conservative amino acid substitutions), an isolated antibody that binds specifically to a peptide cited above, methods of selling and/or licensing single exon probes or microarrays to a customer desiring to measure gene expression, a method of providing human gene expression data by subscription, and a computer-readable storage medium which contains a database having a plurality of records (each record including data on the expression of a single exon probe cited above. The probe, methods and apparatus are useful in gene expression analysis. The probes may be used as tools for surveying tissues to detect the presence of expressed messages that contain their specific exon, or in constructing genome-derived single exon microarrays. In addition, the probes are used in identifying and characterising alternative splicing events, in detecting and characterising gross alterations in the genomic locus that includes their exon, in assessing smaller genomic alterations, in priming the synthesis of nucleic acids, or in expressing the ORF-encoded peptide. The present sequence is a human single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030194704
Sequence 598 BP; 123 A; 197 C; 158 G; 120 T; 0 U; 0 Other;

Query Match 97.9%; Score 578.4; DB 12; Length 598;
Best Local Similarity 99.7%; Pred. No. 5.3e-132;
Matches 590; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGGAGGGAGGCTTCAAGAGGAACACTGTGATTGGAGGAGGAGGAGGTTGGAG 60
DB 7 ATGGTGGGAGGCTTGAAGAGGAACACTGTGATTGGAGGAGGAGGAGGTTGGAG 66
QY 61 TGGAGTCCAGCAGGCGCTTCAGAGCTACACAGAGCCCTGCTCCGATCTCCCTAGACAAA 120
DB 67 TGGAGTCCAGCAGGCGCTTCAGAGCTACACAGAGCCCTGCTCCGATCTCCCTAGACAAA 126
QY 121 GTCCAGCGCAGCTGGGCCCCCGAGACCCAGCCTCCGAGGAGCATGCTCATCATTAAC 180
DB 127 GTCCAGCGCAGCTGGGCCCCCGAGACCCAGCCTCCGAGGAGCATGCTCATCATTAAC 186
QY 181 ACCCTCCACAGCTGCAGCTGCTGCTGGGCTCCGCCCCCTGCGCTG-CCCCCGGA 239
DB 187 ACCCTCCACAGCTGCAGCTGCTGCTGGGCTCCGCCCCCTGCGCTGCCCCCGGA 246
QY 240 GCCCTCTTCTTGGGCGAGGAGGATTTCTCCCTGTGCAGCCACCATTTGGCTCTATCCTCAG 299

DB 247 GCCCTCTTCTTGGGCGAGGAGGATTTCTCCCTGTGCAGCCACCATTTGGCTCTATCCTCAG 306
QY 300 GGAGCTGGACACCTTCATGGATGGGACTGAGCCCCCTCAGATCCAGTGCATCCCTCTGG 359
DB 307 GGAGCTGGACACCTTCATGGATGGGACTGAGCCCCCTCAGATCCAGTGCATCCCTCTGG 366
QY 360 CTCCAGAAATGAAGTGGCCACCCAGCCTGATCCAGTCTTTCTTAGAAGCTCTTGAGCTCCCG 419
DB 367 CTCCAGAAATGAAGTGGCCACCCAGCCTGATCCAGTCTTTCTTAGAAGCTCTTGAGCTCCCG 426
QY 420 GTACTTGGGGGAGCTCTGGCCTGGATGACTTTCTTTTGGCAATGACACATCTGCGGTAGA 479
DB 427 GTACTTGGGGGAGCTCTGGCCTGGATGACTTTCTTTCTGGACATTTGACACATCTGCGGTAGA 486
QY 480 AAGGAGCTTCGACGGGCCCCCAGCAGAGCCTCTTCACAACTCTTCTGCCCCCAGGTTTC 539
DB 487 AAGGAGCTTCGACGGGCCCCCAGCAGAGCCTCTTCACAACTCTTCTGCCCCCAGGTTTC 546
QY 540 TTGGGAGTGGAAATGAATGGAATCATCATCATGGAATCATTTCTGGGCTCTTAA 591
DB 547 TTGGGAGTGGAAATGAATGGAATCATCATCATGGAATCATTTCTGGGCTCTTAA 598

RESULT 11
AAH34147
ID AAH34147 standard; cDNA; 1596 BP.
XX AAH34147;
AC AAH34147;
DT 03-SEP-2001 (first entry)
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1229.
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1229.
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
OS Homo sapiens.
XX WO200122920-A2.
FN 05-APR-2001.
PD 28-SEP-2000; 2000WO-US026524.
XX 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
XX P-PSDB; AAG74742.
DR Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
PT Claim 1; Page 3003-3004; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients' own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal

CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 1596 BP; 354 A; 418 C; 410 G; 409 T; 0 U; 5 Other;

Query Match 97.5%; Score 576.2; DB 4; Length 1596;
Best Local Similarity 99.3%; Pred. No. 2.3e-131;
Matches 587; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

QY 1 ATGGAGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGAGGAGG 60
Db |||||||
QY 345 ATGGAGGAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGAGGAG 404
Db |||||||

QY 61 TGGAGTCCAGAGGCTTCCAGAGCTACCAAGAGCCCTGCTCCGATCTCCCTAGACAAA 120
Db |||||||

QY 405 TGGAGTCCAGAGGCTTCCAGAGCTACCAAGAGCCCTGCTCCGATCTCCCTAGACAAA 464
Db |||||||

QY 121 GTCCAGCGAGCTGGGCCCCGAGCACCCAGCCCTCCGAGGCATGTCTCATCATTAAC 180
Db |||||||

QY 465 GTCCAGGC - CTTGGGCCCCGAGCACCCAGCCCTCCGAGGCATGTCTCATCATTAAC 522
Db |||||||

QY 181 ACCCTCCAGAGCTGAGGCTGCACTTGGCTGGCTCCGCCCCCTGCTCCGCCCCGAG 240
Db |||||||

QY 523 ACCCTCCAGAGCTGAGGCTGCACTTGGCTGGCTCCGCCCCCTGCTCCGCCCCGAG 582
Db |||||||

QY 241 CCCCTCTTCTGGGAGGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 300
Db |||||||

QY 583 CCCCTCTTCTGGGAGGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAGG 642
Db |||||||

QY 301 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 360
Db |||||||

QY 643 GAGCTGGACACCTCCATGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 702
Db |||||||

QY 361 CTCAGATGAAGTGCACCCAGCCCTGATCAGATCTTCTTGAAGCTCTGAGTCCCGG 420
Db |||||||

QY 703 CTCAGATGAAGTGCACCCAGCCCTGATCAGATCTTCTTGAAGCTCTGAGTCCCGG 762
Db |||||||

QY 421 TACTTGGGGGACTCTGGCTGTGATGACTTCTTCTTGGACATTCGACATTCGCGGTAGAA 480
Db |||||||

QY 763 TACTTGGGGGACTCTGGCTGTGATGACTTCTTCTTGGACATTCGACATTCGCGGTAGAA 822
Db |||||||

QY 481 AAGGAGCTTGCAGGCCCCCAGAGCCCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 540
Db |||||||

QY 823 AAGGAGCTTGCAGGCCCCCAGAGCCCTCTCTCAACCTCTTCTGTGCCCCAGGTTCT 591
Db |||||||

QY 541 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
Db |||||||

QY 883 TGGGAGTGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933
Db |||||||

RESULT 12
ABL83086
ID ABL83086 standard; cDNA; 506 BP.
XX
AC ABL83086;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:6064.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
FN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US017756.
XX

PR 26-MAY-2000; 2000US-0207484P.
XX (CORI-) CORIXA CORP.
PI Algate PA, Harlocker SL, Jones R;
XX WPI; 2002-122075/16.
DR
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
PS Claim 1; SEQ ID NO 6064; 489pp; English.
XX
CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 506 BP; 98 A; 184 C; 124 G; 100 T; 0 U; 0 Other;

Query Match 80.0%; Score 473; DB 6; Length 506;
Best Local Similarity 98.6%; Pred. No. 3.6e-106;
Matches 488; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 60 GTGGAGTCCAGAGGCTTTCAGAGTACAGAGAGCCCTGCTCCGATCTCCCTAGACAA 119
Db |||||||

QY 14 GCGGAGTCCAGAGGCTTTCAGAGTACAGAGAGCCCTGCTCCGATCTCCCTAGACAA 73
Db |||||||

QY 120 AGTCCAGAGGAGGCTTGGGCCCCGAGAGCCAGCCCTCGAGGCATGTCTCATCATAA 179
Db |||||||

QY 74 AGTCCAGGCG - CCTGGGCCCCGAGAGCCAGCCCTCGAGGCATGTCTCATCATAA 131
Db |||||||

QY 180 CACCCCTCCAGAGCTGAGGCTGCACTTTCGCTGGCTCCCGCCCTGCGCCCTCCCGG 239
Db |||||||

QY 132 CACCCCTCCAGAGCTGAGGATGCACTTTCGCTGGCTCCCGCCCTGCGCCCTCCCGG 191
Db |||||||

QY 240 GCCCTCTTCTTGGGAGGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAG 299
Db |||||||

QY 192 GCCCTCTTCTTGGGAGGAGGATTTCTCCCTGTGAGCCACCATTTGGCTCTATCTCAG 251
Db |||||||

QY 300 GGAGCTGGACACCTCCATGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 359
Db |||||||

QY 252 GGAGCTGGACACCTCCATGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 311
Db |||||||

QY 360 CTTCCAGAGTGAAGTGGACACCCAGCCCTGATCCAGTCTTCTTGAAGCTCTGAGTCCCG 419
Db |||||||

QY 312 CTTCCAGAGTGAAGTGGACACCCAGCCCTGATCCAGTCTTCTTGAAGCTCTGAGTCCCG 371
Db |||||||

QY 420 GTACTTGGGGGACTCTGGCTGGATGACTTCTTCTTGGACATTTGACATCTGCGGTAGA 479
Db |||||||

QY 372 GTACTTGGGGGACTCTGGCTGGATGACTTCTTCTTGGACATTTGACATCTGCGGTAGA 431
Db |||||||

QY 480 AAGGAGCTTGCAGGCCCCCAGAGCCCTCTCAACCTCTTCTGTGCCCCAGGTTTC 539
Db |||||||

QY 432 AAGGAGCTTGCAGGCCCCCAGAGCCCTCTCAACCTCTTCTGTGCCCCAGGTTTC 491
Db |||||||

```
QY 540 TTGGGAGTGAATGA 554
Db 492 TTGGGAGTGAATGA 506

RESULT 13
ABL83261
ID ABL83261 standard; cDNA; 205 BP.
XX
AC ABL83261;
DT 17-MAY-2002 (first entry)
DE Human ovarian cancer related cDNA clone SEQ ID NO:6239.
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
PN WO200192581-A2.
XX
PD 06-DEC-2001.
PF 29-MAY-2001; 2001WO-US017756.
PR 26-MAY-2000; 2000US-0207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
PS Claim 1; SEQ ID NO 6239; 489pp; English.
XX
CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABU77023 to ABL87934.
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 205 BP; 36 A; 89 C; 45 G; 35 T; 0 U; 0 Other;
Query Match 31.6%; Score 186.6; DB 6; Length 205;
Best Local Similarity 97.1%; Pred. No. 5.2e-36;
Matches 201; Conservative 0; Mismatches 4; Indels 2; Gaps 1;
QY 63 GAGTCACAGAGCCTTCAGAGTACAGCAAGCCTGCTCCGATCTCCCTAGACAAAGT 122
Db 1 GAGTCACAGAGCCTTCAGAGTACAGCAAGCCTGCTCCGATCTCCCTAGACAAAGT 60
QY 123 CCAGCGCAGCCTGGGCCCCCGAGCAGCCAGCCTCCGAGGATGTCTCATCAACAC 182
```

```
Db 61 CCAGCGC--CTGGGCCCGCCGAGCAGCCAGCAGCTCCGAGCATTTGCTCTCATCAACAC 118
QY 183 CTTCCACAGAGTGCAGGCTGCACCTTCGCTCGCTCCCGCCCTGCGCCCTCCCGAGCC 242
Db 119 CTTCCACAGAGTGCAGGCTGCACCTTCGCTCGCTCCCGCCCTGCGCCCTCCCGAGCC 178
QY 243 CTTCTTCTCTGGGCGGAGGAGGATTTCTC 269
Db 179 CTTCTTCTCTGGGCGGAGGAGGATTTCTC 205

RESULT 14
ACH77983
ID ACH77983 standard; DNA; 526 BP.
XX
AC ACH77983;
DT 29-JUL-2004 (first entry)
DE Human genome derived single exon probe #11178.
XX
KW Human; probe; ss; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX
OS Homo sapiens.
PN US2003194704-A1.
XX
PD 16-OCT-2003.
XX
PF 03-APR-2002; 2002US-00029386.
XX
PR 03-APR-2002; 2002US-00029386.
XX
PA (PENN/) PENN S G.
PA (RANK/) RANK D R.
PA (HANZ/) HANZEL D K.
XX
PI Penn SG, Rank DR, Hanzel DK;
XX
DR WPI; 2004-119264/12.
XX
PT New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.
XX
PS Claim 15; SEQ ID NO 11178; 80pp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
```

expression analysis. The probes may be used as tools for surveying tissues to detect the presence of expressed messages that contain their specific exon, or in constructing genome-derived single exon microarrays. In addition, the probes are used in identifying and characterising alternative splicing events, in detecting and characterising gross alterations in the genomic locus that includes their exon, in assessing smaller genomic alterations, in priming the synthesis of nucleic acids, or in expressing the ORF-encoded peptide. The present sequence is a human single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030194704

Sequence 526 BP; 105 A; 133 C; 134 G; 154 T; 0 U; 0 Other;

Query Match 28.1%; Score 166; DB 12; Length 526;
Best Local Similarity 100.0%; Pred. No. 7.4e-31;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 GGGGACTCTGCGCTGGATGACTTCTTTCTGGACATTTGACATTCGCGGTAGAAAAGGA 485
DB 1 GGGGACTCTGCGCTGGATGACTTCTTTCTGGACATTTGACATTCGCGGTAGAAAAGGA 60

QY 486 GCCTCAGCGGCCCCCACCAGAGCCCTCCTCACACCTTCTCTGCCCGCTTCTGGGA 545
DB 61 GCCTCAGCGGCCCCCACCAGAGCCCTCCTCACACCTTCTCTGCCCGCTTCTGGGA 120

QY 546 GTGGGAATGAATCGGATCATCATCATGGAATCATTTCTGGGGTCTTAA 591
DB 121 GTGGGAATGAATCGGATCATCATCATGGAATCATTTCTGGGGTCTTAA 166

RESULT 15

AAL41497
ID AAL41497 standard; DNA; 2082 BP.

AC AAL41497;

DT 19-DEC-2002 (first entry)

DE Mouse haematopoietic progenitor protein (Hepp) gene.

XX Neuroprotective; nontropic; cytostatic; neurodegenerative disease; blood;
KW amyotrophic sclerosis; haematological disorder; neoplasm; leukaemia;
KW acute myelomonocytic leukaemia; lymphoblastic lymphoma; multiple myeloma;
KW chronic lymphocytic leukaemia; acute lymphoblastic leukaemia;
KW B-prolymphocytic leukaemia; plasma cell leukaemia; large B-cell lymphoma;
KW adult T-cell lymphoma; nodal marginal zone B-cell lymphoma; stem cell;
KW Burkitt's lymphoma; follicular lymphoma; hairy cell leukaemia;
KW mantle cell lymphoma; splenic marginal zone B-cell lymphoma;
KW T-prolymphocytic leukaemia; haematopoietic cytokine; growth factor;
KW progenitor cell; gene therapy; Hepp; haematopoietic progenitor protein;
KW mouse; gene; ds.

OS Mus musculus.

XX Key Location/Qualifiers
FT CDS 191..904

FT /*tag= a
FT /product= "Mouse Hepp protein"
FT 2045..2049
FT /*tag= b

XX WO200266610-A2.

XX 29-AUG-2002.

XX 15-FEB-2002; 2002WO-US004503.

XX 16-FEB-2001; 2001US-0268923P.

XX (UWMI-) UNIV MIAMI.

XX

PI Juresic R, Nachtmann RG;

XX WPI; 2002-674928/72.

DR P-PSDB; AAO22897.

XX New hematopoietic progenitor protein (Hepp) genes and proteins, useful

PT for detecting, treating and preventing neurodegenerative diseases, e.g.
PT amyotrophic sclerosis, and hematological disorders, e.g. neoplasms of the
PT blood.

PS Claim 5; Fig 1A; 54pp; English.

XX The invention relates to an isolated nucleic acid comprising at least 85%
CC identity to either of 2 2082 base pair sequences, given in the
CC specification. The nucleic acids and polypeptides of the invention are
CC useful for detecting, treating and preventing neurodegenerative diseases
CC such as amyotrophic sclerosis, and hematological disorders, particularly
CC neoplasms of the blood such as acute myelomonocytic leukaemia,
CC lymphoblastic lymphoma, chronic lymphocytic leukaemia, acute
CC lymphoblastic leukaemia, multiple myeloma, B-prolymphocytic leukaemia,
CC plasma cell leukaemia, adult T-cell lymphoma/leukaemia, diffuse large B-
CC cell lymphoma, nodal marginal zone B-cell lymphoma, Burkitt's lymphoma,
CC follicular lymphoma, hairy cell leukaemia, mantle cell lymphoma, splenic
CC marginal zone B-cell lymphoma, and T-prolymphocytic leukaemia. They are
CC also useful as reagents for differential identification of tissues and
CC cell types present in the biological sample. The mammal is useful in
CC screening drugs for treating the disorders cited above, and for testing
CC of novel haematopoietic cytokines/growth factors for mobilisation and
CC differentiation of stem and progenitor cells. The nucleic acids of the
CC invention can be used in gene therapy. This polynucleotide sequence
CC represents the mouse haematopoietic progenitor protein (Hepp) gene of the
CC invention

XX Sequence 2082 BP; 484 A; 520 C; 522 G; 556 T; 0 U; 0 Other;

Query Match 8.0%; Score 47.4; DB 6; Length 2082;

Best Local Similarity 55.0%; Pred. No. 0.13;

Matches 115; Conservative 0; Mismatches 91; Indels 3; Gaps 1;

QY 8 GAGGCTTGAAGAGGAAACACTCTGATTTGGAGAGGAGGAGGAGGTGGAGTGGAGTC 67

DB 201 GAGGCTTGAAGAGGAAATATGTTGACAGAGAGAGGAGTAGAGGGTTTG---GCACGTG 257

QY 68 CAGCAGGCTTTCAGAGCTACCAAGAGCCCTGCTCCGATCTCCCTAGACAAGTCCAGC 127

DB 258 TCCCTTCTATAGCTGCGAGCGACAGTCACCTCTGGACATGTCCCTTGTCAAGCTCCAGC 317

QY 128 GCAGCTGGGCCCCCGAGCAGCCAGCCCTCCGAGGCATGTCTCATCATACACCCCTCC 187

DB 318 TCTGTACATGCTAGTGGAGGCCCAATCTCTGCGCTCGGTCTCTATCGCCACACAGTCC 377

QY 188 AACAGCTGCAGGCTGCATCTTGCCTGGCT 216

DB 378 GGCAGATCAGAGGAAATGAGCCAGGAT 406

Search completed: April 26, 2005, 08:27:30

Job time : 429 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2005, 06:54:54 ; Search time 155 Seconds
(without alignments)
6238.965 Million cell updates/sec

Title: US-10-069-386A-1

Perfect score: 591
Sequence: 1 atgagaggagcttgagag.....aaatcattctgggtctctaa 591

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813859 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.2	8.3	7218	1	US-08-232-463-14
2	46.4	7.9	44952	4	US-09-949-016-12197
3	46.4	7.9	44960	4	US-09-949-016-17583
4	46.4	7.9	51552	4	US-09-733-294A-30
5	42.6	7.2	1603	4	US-10-000-489-71
6	41	6.9	1926	3	US-09-249-585A-2
7	41	6.9	1926	4	US-09-410-393-3
8	41	6.9	2580	3	US-09-050-863-2
9	41	6.9	2580	3	US-09-359-081-2
10	41	6.9	5452	2	US-09-130-114-1
11	41	6.9	8705	4	US-09-647-344A-14
12	41	6.9	9600	3	US-08-910-647-1
13	41	6.9	9596	1	US-07-884-811-15
14	41	6.9	10596	1	US-07-885-971-15
15	41	6.9	10596	1	US-08-087-783A-15
16	41	6.9	10596	1	US-08-194-088B-15
17	41	6.9	10596	2	US-08-194-087-15
18	41	6.9	10596	2	PCT-US93-04648-15
19	41	6.9	16080	5	US-09-724-566A-48
20	41	6.9	16080	4	US-09-471-669A-48
21	41	6.9	50937	3	US-09-428-517-1
22	40.8	6.9	396	1	US-07-872-678A-9
23	40.6	6.9	2097	3	US-08-941-445A-10
24	40.4	6.8	2254	4	US-09-016-434-1126
25	40.2	6.8	2573	2	US-08-884-681-2
26	40.2	6.8	2573	3	US-09-258-643-2
27	40.2	6.8	2573	3	US-09-258-643-2

28	40.2	6.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
29	40.2	6.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
30	40	6.8	2127	4	US-09-902-540-8358	Sequence 8358, Ap
31	40	6.8	7628	4	US-09-902-540-875	Sequence 875, App
32	39.6	6.7	1432	4	US-09-620-312D-420	Sequence 420, App
33	39.6	6.7	2825	3	US-09-196-390-5	Sequence 5, Appli
34	39.6	6.7	2825	4	US-09-952-677-5	Sequence 5, Appli
35	39.6	6.7	4394	1	US-08-095-734-1	Sequence 1, Appli
36	39.6	6.7	4394	2	US-08-444-623-1	Sequence 1, Appli
37	39.6	6.7	4394	3	US-08-471-869-1	Sequence 1, Appli
38	39.6	6.7	4394	3	US-09-342-563-1	Sequence 1, Appli
39	39.6	6.7	4394	5	PCT-US94-08267-1	Sequence 1, Appli
40	39.4	6.7	780	4	US-09-902-540-4017	Sequence 4017, Ap
41	39.4	6.7	2481	4	US-09-894-998A-35	Sequence 35, Appl
42	39.4	6.7	2481	4	US-10-237-551-35	Sequence 35, Appl
43	39.4	6.7	3066	4	US-10-237-551-152	Sequence 152, App
44	39.4	6.7	26533	4	US-09-902-540-1199	Sequence 1199, Ap
45	39.4	6.7	154746	4	US-09-827-688-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F18
US-08-232-463-14

Query Match 8.3%; Score 49.2; DB 1; Length 7218;

	: LOCATION: (35942)...(36013)	
	: OTHER INFORMATION: exon 10	
	: NAME/KEY: intron	
	: LOCATION: (36014)...(37884)	
	: OTHER INFORMATION: intron 10	
	: NAME/KEY: exon	
	: LOCATION: (37885)...(38073)	
	: OTHER INFORMATION: exon 11	
	: NAME/KEY: intron	
	: LOCATION: (38074)...(41874)	
	: OTHER INFORMATION: intron 11	
	: NAME/KEY: exon	
	: LOCATION: (41875)...(42001)	
	: OTHER INFORMATION: exon 12	
	: NAME/KEY: intron	
	: LOCATION: (42002)...(42881)	
	: OTHER INFORMATION: intron 12	
	: NAME/KEY: exon	
	: LOCATION: (42882)...(42943)	
	: OTHER INFORMATION: exon 13	
	: NAME/KEY: intron	
	: LOCATION: (42944)...(46129)	
	: OTHER INFORMATION: intron 13	
	: NAME/KEY: exon	
	: LOCATION: (46130)...(46254)	
	: OTHER INFORMATION: exon 14	
	: NAME/KEY: intron	
	: LOCATION: (46255)...(47035)	
	: OTHER INFORMATION: intron 14	
	: NAME/KEY: exon	
	: LOCATION: (47036)...(47173)	
	: OTHER INFORMATION: exon 15	
	: NAME/KEY: intron	
	: LOCATION: (47174)...(47709)	
	: OTHER INFORMATION: intron 15	
	: NAME/KEY: exon	
	: LOCATION: (47710)...(50544)	
	: OTHER INFORMATION: exon 16	
	: US-09-733-294A-30	

Query Match 7.9%; Score 46.4; DB 4; Length 51552;
 Best Local Similarity 50.4%; Pred. No. 0.024;
 Matches 139; Conservative 0; Mismatches 136; Indels 1; Gaps 1;

QY	45	GGAGGAGAGTGGGAGTGAGTCCAGCAGGCCTTCAGAGCTACCAAGCGCCCTGCTCCG	104
DB	48820	GGAGGAGGAGGAGGGGGGCCAGCCAGGCTTCCCCTATCTTCCCGCCACCAGACGAG	48761
QY	105	CATCTCCCTTAGCAAAGTCCAGCAGCAGCCTGGGGCCCCCGAGCACCCAGCCTCCGACGCA	164
DB	48760	CCCCCCCCCATCACCCTGCGCACCCAGCTGGGGCCCCCATCATCCTGCCACCTGGCCA	48701
QY	165	TGTCTCATCATACACCTT-CGAACAGCTGCAGGCTGCACTTGCCTGGCTCCCGCCC	223
DB	48700	GSCCTTCCATCATCCCGCGGCCAAAGTGGGGCCCCAGCATCCCTGTGCTTCCGGC	48641
QY	224	CTGCCTGCCCCCGAGCCCCCTTCTTCTGGGGGAGGAGATTCTCCCTGTCCAGCCACA	283
DB	48640	CTGGACTTAAGTTATGTCTTCAGGGTGGGGGCTCCCACTGTCTATCCCCTACCTCCT	48591
QY	284	TTGGCTCTATCTCAGGAGCTGGACACTCCATGG	319
DB	48580	TCCCTCTCTGCTCAGCATCAGAAACCTCCAGG	48545

RESULT 5
 US-10-000-489-71
 : Sequence 71, Application US/10000489
 : Patent No. 6794363
 : GENERAL INFORMATION:
 : APPLICANT: Benjanin, Stephane
 : APPLICANT: Tanaka, Hiroaki
 : TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

```

; FILE REFERENCE: 9A.US6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 71
; LENGTH: 1603
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..7
; NAME/KEY: CDS
; LOCATION: 8..763
; NAME/KEY: 3'UTR
; LOCATION: 764..1603
; NAME/KEY: polyA_signal
; LOCATION: 1562..1567
; NAME/KEY: polyA_site
; LOCATION: 1588..1603
; US-10-000-489-71

```

	Query Match	7.2%	Score 42.6;	DB 4;	Length 1603;
	Best Local Similarity	54.0%;	Pred. No. 0.066;		
	Matches 87;	Conservative 0;	Mismatches 74;	Indels 0;	Gaps 0;
Qy	131	GCCTGGGCCCCCGAGGACCCAGCCTCCGCGAGGCGATGCTCATCCATAACACACCTCCAAAC	190		
Db	870	GCACGGGAACCCCGCCCACTCCCAATCCCGGGCCCGCCCTCTCCCAACCCGTGCTTCCCC	929		
Qy	191	AGCTGCAGGCTGCACCTTTCGCGTGGCTCCGCGCCCTGCCTGCCCGCCCGAGCCCTCTTCC	250		
Db	930	CGTCCACCCCTCACCTCACCTCGCCCGCCCGGCCACCATCGCGCCCGGGGCTGTATT	989		
Qy	251	TGGCGGAGGAGGATTTTCCTCTGCACCCACCAATGGCTCT	291		
Db	990	TGTTTCGGCTGGGCTTCGGTTCGGGGGCTGTCTCCCTCGGCTCT	1030		

RESULT 6
IIS-09-249-585A-2/c

```

; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE
; FILE REFERENCE: 0867/00905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1926

```

```

; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

```

Query Match	6.9%;	Score 41;	DB 3;	Length 1926;
Best Local Similarity	51.4%;	Pred. No. 0.2;		
Matches	95;	Conservative	0;	Mismatches 90; Indels 0; Gaps 0;
Qy	67	CCAGCAGGCCTTCAGAGCTACCAAGCAAGCCCTCTCCGCATCTCCCTAGACAAAGTCCAG	126	
Db	1063	CCCGCGGCTCCACTACCTCTCGACCCCGGCTCCACTACCTCTCGACCCCGGCCTC	1004	
Qy	127	CGAGCCTGGGCCCCCGAGGACCCAGCCTCGGAGGCATGTCTCATCATACACCCCTC	186	
Db	1003	CACTGCTCTCTGACCCCGGCTCCACCTCTGCTCTGCCCCCTCTGCTCTGCCCCCTC	944	
Qy	187	CAACAGCTGAGGCTGCACCTTGCGCTGGCTCCCGCCCCCTGCGCCCTGCCCCGAGCCCCCTC	246	
Db	943	CTCTGCTCTGCCCCCTCTGCCCCCTCTGCTCTGCCCCCTCTGCCCCCTCTGCTCTGCT	884	
Qy	247	TTCCT	251	
Db	883	CCCT	879	

RESULT 7
US-09-410-399-3/c

```

: Sequence 3, Application US/09410399
: Patent No. 6482587
: GENERAL INFORMATION:
: APPLICANT: Robertson, Erle S.
: APPLICANT: Cotter, Murray A.
: TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
: TITLE OF INVENTION: to Genomic Host DNA
: FILE REFERENCE: UM-03778
: CURRENT APPLICATION NUMBER: US/09/410,399
: CURRENT FILING DATE: 1999-10-01
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 1926
: TYPE: DNA
: ORGANISM: Epstein-Barr virus
: US-09-410-399-3

```

Query Match	6.9%;	Score 41;	DB 4;	Length 1926;
Best Local Similarity	51.4%;	Pred. No. 0.2;		
Matches	95;	Conservative	0;	Mismatches 90;
			Indels	0;
			Gaps	0;
Qy	67	CCAGCAGGCGCTTACGAGCTACCAAGCAAGCGCTCTCCGCATCTCCCTAGACAAAGTCCAG	126	
Db	1063	CCGCGGCGCTCCACTACCTCTCGACCCCGGCTCCACTTACCTCTCGACCCCGGCTC	1004	
Qy	127	CGAGCTGTGGCCCCCGAGGACCCAGCGCTCGGAGGCGATGTCTCTATCATACACGCTC	186	
Db	1003	CACGTGCTCTCGACCCCGGCTTCCACCTCTCTGCCCCCTGCTCTCTGCCCCCTC	944	
Qy	187	CAACAGTGCAGGCTGCACCTTGCCCTTGCTGCCGCCCTGCCCCCTGCCCCGAGCCCCCTC	246	
Db	943	CTCCTGTCTCTGCCCCCTCTCTGCCCCCTCTGCTCTGCCCCCTCTGCTGCTGCTG	884	
Qy	247	TTCTT	251	
Db	883	CCCTT	879	

RESULT 8
US-09-050-863-2/c

Sequence 2, Application US/09050863
Patent No. 611411
GENERAL INFORMATION:
APPLICANT: .Lao, ying
APPLICANT: Hiang, Betty
APPLICANT: Payan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
TITLE OF INVENTION: System
NUMBER OF SEQUENCES: 5

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
;; STREET: 4 Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94111-4187
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/050,863
;; FILING DATE: 30-MAR-1998
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Silva, Robin M.
;; REGISTRATION NUMBER: 38,304
;; REFERENCE/DOCKET NUMBER: A-55638/DJB/RMS
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 781-1989
;; TELEFAX: (415) 949-8711
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2580 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 6.9%; Score 41; DB 3; Length 2580;
Best Local Similarity 51.4%; Pred. No. 0.22;
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 67 CCAGCAGGCTTCAGAGTACAGAGCCCTGCTCCGATCTCCCTAGACAAAGTCCAG 126
Db 1446 CCGCGCGCTCCACTACTCTCTCGACCCCGGCTCCACTACTCTCTCGACCCCGGCTC 1387

QY 127 CGCAGCCTGGGCGGCGGAGCAGCCAGCTCCGAGGATGCTCATCATCAACACCCCTC 186
Db 1386 CACTGCTCTCGACCCGCGCTCCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1327

QY 187 CAACAGCTGCAGGCTGCATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246
Db 1326 CTCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1267

QY 247 TTCCT 251
Db 1266 CCCCT 1262

RESULT 9
US-09-359-081-2/c
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; Hiang, Betty
; Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/359,081
;; FILING DATE: 22-Jul-1999
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/050,863
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Silva, Robin M.
;; REGISTRATION NUMBER: 38,304
;; REFERENCE/DOCKET NUMBER: A-55638/DJB/RMS
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 781-1989
;; TELEFAX: (415) 949-8711
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2580 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match 6.9%; Score 41; DB 3; Length 2580;
Best Local Similarity 51.4%; Pred. No. 0.22;
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 67 CCAGCAGGCTTCAGAGTACAGAGCCCTGCTCCGATCTCCCTAGACAAAGTCCAG 126
Db 1446 CCGCGCGCTCCACTACTCTCTCGACCCCGGCTCCACTACTCTCTCGACCCCGGCTC 1387

QY 127 CGCAGCCTGGGCGGCGGAGCAGCCAGCTCCGAGGATGCTCATCATCAACACCCCTC 186
Db 1386 CACTGCTCTCGACCCGCGCTCCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1327

QY 187 CAACAGCTGCAGGCTGCATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246
Db 1326 CTCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1267

QY 247 TTCCT 251
Db 1266 CCCCT 1262

RESULT 10
US-09-130-114-1
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damej, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Epithomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
US-09-130-114-1

Query Match 6.9%; Score 41; DB 2; Length 5452;
Best Local Similarity 51.4%; Pred. No. 0.31;
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy	67	CCAGCAGCCTT	CGAGCT	ACCGACGAGCCCTGCTCGCATCTCCCTAGACAAGTCCAG	126
Db	1359	CCCGCGGCGCT	TCCACT	ACCTCTCGACCCCGGGCTCCACTACCTCTCTCGACCCCGGGCTC	1418
Qy	127	CGAGCCTGGGCCCC	CGAGAC	CCGAGCCTCCGAGGGATGCTCATCCATAACACCCCTC	186
Db	1419	CATGCGCTCT	CGACCCCGGGCT	CCACTCTGTCTCTGCGCCCTCTGTCTCTGCGCCCTC	1478
Qy	187	CAACAGCTG	CAGGCTG	CACTTGGCTTGGCTTCCGCGCCCTGCGCTTGGCCCCCGGAGCCCTC	246
Db	1479	CTCCTGTCT	GTGCCCTC	CTTGGCCCTCTCTGTCTCTGTGCCCCCTCTGTGCCCCCTCTGTCTCTG	1538
Qy	247	TTCCCT	251		
Db	1539	CCCCCT	1543		

```

RESULT 11
US-09-647-344A-14
; Sequence 14, Application US/09647344A
; Patent No. 6586180
; GENERAL INFORMATION:
; APPLICANT: Rufner, Duane E.
; APPLICANT: Pierce, Michael L.
; APPLICANT: Chen, Zhidong
; TITLE OF INVENTION: Directed Antisense Libraries
; FILE REFERENCE: T6678.PCT.US
; CURRENT APPLICATION NUMBER: US/09/647,344A
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: PCT/US99/06742
; PRIOR FILING DATE: 1999-03-28
; NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 14
; LENGTH: 8705
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pShuttle
US-09-647-344A-14

```

```

RESULT 12
US-08-910-647-1/c
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRES:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street

```

```

RESULT 13
US-09-620-925-1/c
; Sequence 1, Application US/09620925
; Patent No. 6468986
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; Polynucleotide Delivery
;
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/620.925

```

```
;
; FILING DATE: 21-Jul-2000
; CLASSIFICATION: <Unknown>
; PRIORITY INFORMATION:
; APPLICATION NUMBER: 08/910,647
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1

Query Match      6.9%; Score 41; DB 3; Length 9600;
Best Local Similarity 51.4%; Pred. No. 0.39;
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 67 CCAGCAGCGCTTTCAGAGCTACCAAGCCCTGCTCCGCGATCTCCCTAGACAAAGTCCAG 126
DB 1492 CCGGCGGCTCCACTACTCTCTCGACCCGCGCTCCACTACTCTCGACCCGCGCTC 1433
QY 127 CGAGCGTGGGCGGCGGAGCAGCCCTGCTCCGCGATCTCCCTAGACAAAGTCCAG 186
DB 1432 CACTGCTCTGCGGCGGCGGCTCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1373
QY 187 CAACAGCTGCGGCGGCGGAGCAGCCCTGCTCCGCGATCTCCCTAGACAAAGTCCAG 246
DB 1372 CTCCTGCTCTGCGGCGGCGGCTCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1313
QY 247 TTCCT 251
DB 1312 CCCCT 1308

RESULT 14
US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dregler, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID

;
; REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; TOPOLOGY: linear
US-07-884-811-15

Query Match      6.9%; Score 41; DB 1; Length 10596;
Best Local Similarity 51.4%; Pred. No. 0.4;
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 67 CCAGCAGCGCTTTCAGAGCTACCAAGCCCTGCTCCGCGATCTCCCTAGACAAAGTCCAG 126
DB 3026 CCGGCGGCTCCACTACTCTCTCGACCCGCGCTCCACTACTCTCTCGACCCGCGCTC 2967
QY 127 CGAGCGTGGGCGGCGGAGCAGCCCTGCTCCGCGATCTCCCTAGACAAAGTCCAG 186
DB 2966 CACTGCTCTGCGGCGGCGGCTCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2907
QY 187 CAACAGCTGCGGCGGCGGAGCAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246
DB 2906 CTCCTGCTCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2847
QY 247 TTCCT 251
DB 2846 CCCCT 2842

RESULT 15
US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dregler, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2005, 09:13:38 ; Search time 500 Seconds
(without alignments)
7175.983 Million cell updates/sec

Title: US-10-069-386A-1
Perfect score: 591
Sequence: 1 atggaggagcttgaagag.....aaatcattctgggtcctaa 591

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq.*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
6: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq.*
13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
17: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
18: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
19: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
20: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
21: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq.*
22: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	578.4	97.9	598	16	US-10-029-386-24898
2	576.2	97.5	1596	15	US-10-106-698-1239
3	473	80.0	506	9	US-09-867-701-6084
4	186.6	31.6	205	9	US-09-867-701-6239
5	166	28.1	526	16	US-10-029-386-11178
6	47.4	8.0	2082	13	US-10-076-069-1
C 7	46.4	7.9	3179	19	US-10-840-455-20
C 8	46.4	7.9	25138	19	US-10-840-455-44
C 9	46.4	7.9	51552	9	US-09-733-294A-30
C 10	44.8	7.6	440	14	US-10-184-644-202
C 11	44.8	7.6	440	14	US-10-184-634-202

C 12	44.8	7.6	440	16	US-10-063-685-52	Sequence 52, Appl
C 13	44.8	7.6	594	14	US-10-123-155-10	Sequence 10, Appl
C 14	44.8	7.6	594	15	US-10-146-731-10	Sequence 10, Appl
C 15	44.8	7.6	594	15	US-10-140-472-10	Sequence 10, Appl
C 16	44.8	7.6	594	15	US-10-141-761-10	Sequence 10, Appl
C 17	44.8	7.6	594	16	US-10-142-885-10	Sequence 10, Appl
C 18	44.8	7.6	594	16	US-10-158-730-10	Sequence 10, Appl
C 19	44.8	7.6	594	17	US-10-137-871-10	Sequence 10, Appl
C 20	44.8	7.6	594	17	US-10-140-923-10	Sequence 10, Appl
C 21	44.8	7.6	594	17	US-10-141-756-10	Sequence 10, Appl
C 22	44.8	7.6	594	17	US-10-141-759-10	Sequence 10, Appl
C 23	44.8	7.6	594	17	US-10-140-805-10	Sequence 10, Appl
C 24	44.8	7.6	594	17	US-10-140-864-10	Sequence 10, Appl
C 25	44.8	7.6	594	17	US-10-142-426-10	Sequence 10, Appl
C 26	44.8	7.6	2748	18	US-10-437-963-53188	Sequence 53188, A
C 27	42.6	7.2	1603	10	US-09-992-600A-71	Sequence 71, Appl
C 28	42.6	7.2	1603	10	US-09-924-340-71	Sequence 71, Appl
C 29	42.6	7.2	1603	10	US-09-992-095B-71	Sequence 71, Appl
C 30	42.6	7.2	1603	10	US-09-999-570-71	Sequence 71, Appl
C 31	42.6	7.2	1603	14	US-10-000-489-71	Sequence 71, Appl
C 32	42.6	7.2	1603	14	US-10-000-986-71	Sequence 71, Appl
C 33	42.6	7.2	1603	16	US-10-154-678-71	Sequence 71, Appl
C 34	42.6	7.2	1603	16	US-10-001-142-71	Sequence 71, Appl
C 35	42.6	7.2	1603	19	US-10-838-854-71	Sequence 71, Appl
C 36	42.4	7.2	511	16	US-10-029-386-24881	Sequence 24881, A
C 37	42.4	7.2	600	18	US-10-363-345A-39087	Sequence 39087, A
C 38	42.4	7.2	600	18	US-10-363-345A-39088	Sequence 39088, A
C 39	42.4	7.2	600	19	US-10-363-483A-39087	Sequence 39087, A
C 40	42.4	7.2	600	19	US-10-363-483A-39088	Sequence 39088, A
C 41	42.4	7.2	1650	18	US-10-437-963-27296	Sequence 27296, A
C 42	42	7.1	672	18	US-10-437-963-59519	Sequence 59519, A
C 43	42	7.1	2580	15	US-10-156-761-1	Sequence 3895, Ap
C 44	42	7.1	9025608	15	US-10-156-761-1	Sequence 1, Appl
C 45	41.8	7.1	32329	17	US-10-374-903A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-10-029-386-24898
; Sequence 24898, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AROMICA-X-2
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24898
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010271.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.64
; OTHER INFORMATION: EST HUMAN HIT: BG720189.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: O70622, EVALUE 2.30e-01
; OTHER INFORMATION: NT HIT: gi16178011, EVALUE 0.00e+00
US-10-029-386-24898

Query Match 97.9%; Score 578.4; DB 16; Length 598;
Best Local Similarity 99.7%; Pred. No. 38-161;
Matches 590; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGGAGGAGGCTTGAAGAGGAACACTCTGTTTGAAGAGGAGGAGGAGGTGGAG 60

DB 7 ATGGTGGGAGGCTTGAAGAGGAACACTCTGTTTGAAGAGGAGGAGGAGGTGGAG 66

Db 132 CACCCCTCCAAAGCTGCAGATGCACTTCGGCTGGCTCCGGCCCTGCGCTGCCCCCGA 191
Qy 240 GCCCTCTTTCCTGGGCGAGAGATTTCCTGTGTCAGCCACCATTTGGCTCTATCTCAG 299
Db 192 GCCCTCTTTCCTGGGCGAGAGATTTCCTGTGTCAGCCACCATTTGGCTCTATCTCAG 251
Qy 300 GGAGCTGGACACTTCATGATGGAGCTGAGCCCTCAGAACTCAGTCCCTCTGG 359
Db 252 GGAGCTGGACACTTCATGATGGAGCTGAGCCCTCAGAACTCAGTCCCTCTGG 311
Qy 360 CTTCCAGAAATGAAGTGCACCCAGCCTGATCCAGTCTTCTTAGAAGCTCTGAGTCCCG 419
Db 312 CTTCCAGAAATGAAGTGCACCCAGCCTGATCCAGTCTTCTTAGAAGCTCTGAGTCCCG 371
Qy 420 GTACTTGGGGGACTCTGGCTGGATGACTTCTTTCTGGACATTTGACATCTGCGGTAGA 479
Db 372 GTACTTGGGGGACTCTGGCTGGATGACTTCTTTCTGGACATTTGACATCTGCGGTAGA 431
Qy 480 AAAGGAGCTGCACGGGCCCCCAGAGGCTCTCTCAACCTTCTTGTGCCCCAGGTTTC 539
Db 432 AAAGGAGCTGCACGGGCCCCCAGAGGCTCTCTCAACCTTCTTGTGCCCCAGGTTTC 491
Qy 540 TTGGGAGTGGAAATGA 554
Db 492 TTGGGAGTGGAAATGA 506

RESULT 4

US-09-867-701-6239

; Sequence 6239, Application US/09867701

; Patent No. US20020132237A1

; GENERAL INFORMATION:

; APPLICANT: Aglate, Paul A.

; APPLICANT: Jones, Robert

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701

; CURRENT FILING DATE: 2001-05-29

; NUMBER OF SEQ ID NOS: 10912

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6239

; LENGTH: 205

; TYPE: DNA

; ORGANISM: Homo sapien

; US-09-867-701-6239

Query Match

Best Local Similarity 31.6%; Score 186.6; DB 9; Length 205;

Matches 201; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 63 GAGTCGACGAGCTTCAGAGTACCAGCAAGCCCTGCTCCGATCTCCCTAGACAAAGT 122
Db 1 GAGTCGACGAGCTTCAGAGTACCAGCAAGCCCTGCTCCGATCTCCCTAGACAAAGT 60

Qy 123 CCAGGCGACCTGGGCCCCCGAGCAGCCAGCCTCCGAGGATGCTCTCATCCATAACAC 182
Db 61 CCAGGCG--CTGGGCCCCCGAGCAGCCAGCCTCCGAGATGCTCTCATCCATAACAC 118

Qy 183 CTTCCAAAGCTGCAGGCTGCATTGCTGCTGCTCCGCTCCGCTCCGCTCCGCTCCGAGCC 242
Db 119 CTTCCAAAGCTGCAGGCTGCATTGCTGCTGCTCCGCTCCGCTCCGCTCCGCTCCGAGCC 178

Qy 243 CTTCTTCTTGGGAGGAGGATTTCTC 269

Db 179 CTTCTTCTTGGGAGGAGGATTTCTC 205

RESULT 5

US-10-029-386-11178

; Sequence 11178, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

; FILE REFERENCE: AEOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 34288

; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 11178

; LENGTH: 526

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC010271.5

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.64

; OTHER INFORMATION: SWISSPROT HIT: Q60963, EVALUE 2.20e+00

; OTHER INFORMATION: NT HIT: g115718683, EVALUE 0.00e+00

; OTHER INFORMATION: EST_HUMAN HIT: AA426355.1, EVALUE 1.00e-114

; US-10-029-386-11178

Query Match

Best Local Similarity 28.1%; Score 166; DB 16; Length 526;

Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 GGGGACTCTGGCTGGATGACTTCTTCTGACATTTGACATCTGCGGTAGAAAAGGA 485

Db 1 GGGGACTCTGGCTGGATGACTTCTTCTGACATTTGACATCTGCGGTAGAAAAGGA 60

Qy 486 GCCTGACGGGCCCCCAGAGCCTCTTCTGACAACTTCTGCCCCAGGTTCTTGGGA 545

Db 61 GCCTGACGGGCCCCCAGAGCCTCTTCTGACAACTTCTTCTGCCCCAGGTTCTTGGGA 120

Qy 546 GTGGAATGAACTGGATCATCATCATGAAATCATTTCTGGGTCCTAA 591

Db 121 GTGGAATGAACTGGATCATCATCATGAAATCATTTCTGGGTCCTAA 166

RESULT 6

US-10-076-069-1

; Sequence 1, Application US/10076069

; Publication No. US20020177214A1

; GENERAL INFORMATION:

; APPLICANT: JUREGIC, ROLAND

; APPLICANT: NACHTMAN, RONALD

; TITLE OF INVENTION: HEPP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELOP

; FILE REFERENCE: 39532-176599

; CURRENT APPLICATION NUMBER: US/10/076,069

; CURRENT FILING DATE: 2002-02-15

; PRIOR APPLICATION NUMBER: US 60/268,923

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 2082

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (191)..(901)

; US-10-076-069-1

Query Match

Best Local Similarity 8.0%; Score 47.4; DB 13; Length 2082;

Matches 115; Conservative 0; Mismatches 91; Indels 3; Gaps 1;

Qy 8 GAGGCTTGAAGAGGAAACACTCTGATTTGGAAAGAGGAGGAGGTTGGGAGTGGAGTC 67

Db 201 GAGGCTTGAAGAGGAAATATGTTGACCAGAGAGGAGGAGTGGGTTTG---GCACGT 257

Qy 68 CAGCAGGCTTCAGAGCTACCAAGCCCTGCTCCGCATCTCCCTAGACAAGTCCAGC 127

Db 258 TCCCTTCCTATAGCTGCGAGCAGACAGTCACTCTCTGGACATGTCCTTGTCAAGCTCCAGC 317
Qy 128 CGAGCTGGGCCCCGAGCAGCCAGCTCCGAGCATGTCTCTATCATCAATACACCTCC 187
Db 318 TCTGTCACTGCTAGTGAGGCCAATCTCTCGCTCGTCTCTCATCCCAACAGATCC 377
Qy 188 AACAGCTGAGGCTGCACCTTGCCCTGGCT 216
Db 378 GCGATCCAGGAGGAATGAGCCAGGAT 406

RESULT 7
US-10-840-455-20/c
; Sequence 20, Application US/10840455
; Publication No. US20050032094A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Aktiengesellschaft
; APPLICANT: Hagen, Gustav
; APPLICANT: Wick, Marese
; APPLICANT: Zubov, Dmitry
; TITLE OF INVENTION: Regulatory DNA Sequences of the Gene for the Human Catalytic
; TITLE OF INVENTION: Telomerase Subunit, and Their Diagnostic and Therapeutic Use
; FILE REFERENCE: LeA 32 805C1
; CURRENT APPLICATION NUMBER: US/10/840,455
; PRIOR FILING DATE: 2004-05-06
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: US 09/582,246
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: DE19757984.1
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 3179
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-840-455-20

Query Match 7.9%; Score 46.4; DB 19; Length 3179;
Best Local Similarity 50.4%; Pred. No. 0.0019;
Matches 139; Conservative 0; Mismatches 136; Indels 1; Gaps 1;

Qy 45 GGAGGAGGTGGGAGTGAGTCCAGCAGGCTTTCAGAGTACAGCAAGCCTCTCTCG 104
Db 447 GGAGGAGGTGGGAGTGAGTCCAGCAGGCTTTCAGAGTACAGCAAGCCTCTCTCG 388

Qy 105 CATCTCCCTAGACAAAGTCCAGCGAGCTGGGGCCCCGAGCAGCCAGCCTCCGAGGCA 164
Db 387 CCCCCCATCACCTCGCCACCCAGCTGGGGCCCCCATCATCTCTGCCACCTGGCCA 328

Qy 165 TGTCTCATTCATACACCT-CCAAACAGTGCAGGCTGCATCTCGCTGGCTCCCGCC 223
Db 327 GGCCTTCATATCCCGCGCCAAAGCTGGGGCCCCAGCATCTCTGTCTGCTTCCGGC 268

Qy 224 CTGCCCTGCCCCAGCCCTTCTCTGGGCGAGGAGGATTTCTCCCTGTGCAGCACA 283
Db 267 CTGGACTTACTGTTATGTTTCAGGGTGGGGCTCCCACTGTCTATCCCTTACCTCT 208

Qy 284 TTGGCTCTATCTCCTCAGGAGCTGGACACCTCCATGG 319
Db 207 TCCCTCTCTGCTCACAGATCAGAACTCCAGG 172

RESULT 8
US-10-840-455-44/c
; Sequence 44, Application US/10840455
; Publication No. US20050032094A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Aktiengesellschaft
; APPLICANT: Hagen, Gustav
; APPLICANT: Wick, Marese

; APPLICANT: Zubov, Dmitry
; TITLE OF INVENTION: Regulatory DNA Sequences of the Gene for the Human Catalytic
; TITLE OF INVENTION: Telomerase Subunit, and Their Diagnostic and Therapeutic Use
; FILE REFERENCE: LeA 32 805C1
; CURRENT APPLICATION NUMBER: US/10/840,455
; CURRENT FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: PCT/EP98/08216
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: US 09/582,246
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: DE19757984.1
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44
; LENGTH: 25138
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(25138)
; OTHER INFORMATION: n is a, c, g, or t.
US-10-840-455-44

Query Match 7.9%; Score 46.4; DB 19; Length 25138;
Best Local Similarity 50.4%; Pred. No. 0.0025;
Matches 139; Conservative 0; Mismatches 136; Indels 1; Gaps 1;

Qy 45 GGAGGAGGTGGGAGTGAGTCCAGCAGGCTTTCAGAGTACAGCAAGCCTCTCTCG 104
Db 22406 GGAGGAGGTGGGAGTGAGTCCAGCAGGCTTTCAGAGTACAGCAAGCCTCTCTCG 22347

Qy 105 CATCTCCCTAGACAAAGTCCAGCGAGCTGGGGCCCCGAGCAGCCAGCCTCCGAGGCA 164
Db 22346 CCCCCCATCACCTCGCCACCCAGCTGGGGCCCCCATCATCTCTGCCACCTGGCCA 22287

Qy 165 TGTCTCATTCATACACCT-CCAAACAGTGCAGGCTGCATCTCGCTGGCTCCCGCC 223
Db 22286 GGCCTTCATATCCCGCGCCAAAGCTGGGGCCCCAGCATCTCTGTCTGCTTCCGGC 22227

Qy 224 CTGCCCTGCCCCAGCCCTTCTCTGGGCGAGGAGGATTTCTCCCTGTGCAGCACA 283
Db 22226 CTGGACTTACTGTTATGTTTCAGGGTGGGGCTCCCACTGTCTATCCCTTACCTCT 22167

Qy 284 TTGGCTCTATCTCCTCAGGAGCTGGACACCTCCATGG 319
Db 22166 TCCCTCTCTGCTCACAGATCAGAACTCCAGG 22131

RESULT 9
US-09-733-294A-30/c
; Sequence 30, Application US/09733294A
; Patent No. US2002004558A1
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: William Gaarde
; APPLICANT: Susan M. Freier
; APPLICANT: Edward V. Wanciewicz
; TITLE OF INVENTION: ANTISENSE MODULATION OF TERT EXPRESSION
; FILE REFERENCE: ISPH-0527
; CURRENT APPLICATION NUMBER: US/09/733,294A
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 09/572,423
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 108
; SEQ ID NO 30
; LENGTH: 51552
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)...(11492)
; OTHER INFORMATION: exon 1


```
; ORGANISM: Homo Sapien
US-10-184-644-202

Query Match
Best Local Similarity 18.8%; Score 44.8; DB 14; Length 440;
Matches 58; Conservative 81; Mismatches 169; Indels 0; Gaps 0;

QY 127 GCAGCTGGGCGCCCGAGCAGCCAGCTCCGAGGAGGATTTCTCCCTGTCAACCACTTGGCTCTATCTCCTCAGGAGCTG 306
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
389 CSCWSSC.CYRHCSCC..YNC.MS..YKNSBSC.CY.C..CS.CSCT.NC.M.GC.MD 330

QY 187 CAACAGCTGCAGGCTGCATTTGGCTGGCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCC 246
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
329 CNCCSCCDNCSNCSSCASSCWSS.SCSHCYSCSNCSSCSCSSCSCSSCSCSSCSCSSCSCSSCSCSSCSCSSC 270

QY 247 TTCCTGGGCGAGGAGGATTTCTCCCTGTCAACCACTTGGCTCTATCTCCTCAGGAGCTG 306
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
269 SSSCNNHNCSNCSSCS.SCSGCCSNSSSCSSCS...NAGC.N.NSTYBSCRC..TBTC 210

QY 307 GACACCTCCATGGATGGAGCTGAGCCCTCAGAAATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAG 366
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
209 .ANAC.N..CNCC.CW.TC..NKC.SCTSNCRDBW.AC.CC.NC.CC.CC.CC.SC.. 150

QY 367 AATGAAGTGCCACCCAGCTGATCCAGTCTCTTTAGAGCTCTGAGCTCCCGGTACTTG 426
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
149 CDCCSA.WTCSDC.BC.WSCYBTHTCDY.BH.T.YC..DCANC.TDIT.CBYNC.THTTC 90

QY 427 GGGGACTC 434
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
89 .C.B.YBC 82

RESULT 12
US-10-063-685-52/c
; Sequence 52, Application US/10063685
; Publication No. US20030180909A1
; GENERAL INFORMATION:
; APPLICANT: Eaton,Dan L.
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,685
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 52
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-685-52

Query Match
Best Local Similarity 18.8%; Score 44.8; DB 16; Length 440;
Matches 58; Conservative 81; Mismatches 169; Indels 0; Gaps 0;

QY 127 GCAGCTGGGCGCCCGAGCAGCCAGCTCCGAGGAGGATTTCTCCCTGTCAACCACTTGGCTCTATCTCCTCAG 186
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
389 CSCWSSC.CYRHCSCC..YNC.MS..YKNSBSC.CY.C..CS.CSCT.NC.M.GC.MD 330

QY 187 CAACAGCTGCAGGCTGCATTTGGCTGGCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCC 246
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
329 CNCCSCCDNCSNCSSCASSCWSS.SCSHCYSCSNCSSCSCSSCSCSSCSCSSCSCSSCSCSSCSCSSCSCSSC 270

QY 247 TTCCTGGGCGAGGAGGATTTCTCCCTGTCAACCACTTGGCTCTATCTCCTCAGGAGCTG 306
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
269 SSSCNNHNCSNCSSCS.SCSGCCSNSSSCSSCS...NAGC.N.NSTYBSCRC..TBTC 210

QY 307 GACACCTCCATGGATGGAGCTGAGCCCTCAGAAATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAG 366
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
209 .ANAC.N..CNCC.CW.TC..NKC.SCTSNCRDBW.AC.CC.NC.CC.CC.CC.SC.. 150

QY 367 AATGAAGTGCCACCCAGCTGATCCAGTCTCTTTAGAGCTCTGAGCTCCCGGTACTTG 426
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
149 CDCCSA.WTCSDC.BC.WSCYBTHTCDY.BH.T.YC..DCANC.TDIT.CBYNC.THTTC 90

QY 427 GGGGACTC 434
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
89 .C.B.YBC 82

RESULT 11
US-10-184-634-202/c
; Sequence 202, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 202
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-202

Query Match
Best Local Similarity 18.8%; Score 44.8; DB 14; Length 440;
Matches 58; Conservative 81; Mismatches 169; Indels 0; Gaps 0;

QY 127 GCAGCTGGGCGCCCGAGCAGCCAGCTCCGAGGAGGATTTCTCCCTGTCAACCACTTGGCTCTATCTCCTCAG 186
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
389 CSCWSSC.CYRHCSCC..YNC.MS..YKNSBSC.CY.C..CS.CSCT.NC.M.GC.MD 330

QY 187 CAACAGCTGCAGGCTGCATTTGGCTGGCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCCGCCCTCC 246
Db |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
329 CNCCSCCDNCSNCSSCASSCWSS.SCSHCYSCSNCSSCSCSSCSCSSCSCSSCSCSSCSCSSCSCSSCSCSSC 270
```

:	APPLICANT:	Filvaroff,Ellen
:	APPLICANT:	Gao,Wei-Qiang
:	APPLICANT:	Gerritsen,Mary E.
:	APPLICANT:	Goddard,Audrey
:	APPLICANT:	Godowski,Paul J.
:	APPLICANT:	Gurney,Austin L.
:	APPLICANT:	Sherwood,Steven
:	APPLICANT:	Smith,Victoria
:	APPLICANT:	Stewart,Timothy A.
:	APPLICANT:	Tumas,Daniel
:	APPLICANT:	Watanabe,Colin K
:	APPLICANT:	Wood,William
:	APPLICANT:	Zhang,Zemin
:	TITLE OF INVENTION:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
:	TITLE OF INVENTION:	ACIDS ENCODING THE SAME
:	FILE REFERENCE:	P333OR1C323
:	CURRENT APPLICATION NUMBER:	US/10/146,731
:	CURRENT FILING DATE:	2002-05-15
:	Prior Application removed - See File Wrapper or Palm	
:	NUMBER OF SEQ ID NOS:	550
:	SEQ ID NO 10	
:	LENGTH:	594
:	TYPE:	PRT
:	ORGANISM:	Homo Sapien
:	US-10-146-731-10	

Query Match	7.6%;	Score 44.8;	DB 15;	Length 594;
Best Local Similarity	8.7%;	Pred. No. 0.0044;		
Matches	28;	Conservative 135;	Mismatches 160;	Indels 0; Gaps 0

Qy	79	CAGAGCTACGAGCAAGCCCTGGTCCGCATCTCCCTAGACAAGTCCACGCCGCGCTGGGC	138
Dd	556	MM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYS.S.S.SDYSYA.SYSYS.S.SWS	497
Qy	139	CCCCGACGCCAGCCCTCCGAGCATGTCTCTATTCATACACCCCTCCACAGCTGCAG	198
Dd	496	SYSSYSDDDY.CYCCYRYHCSDSYSYSYY.CRCCYYT.SYSRYDCHYSCCSDDYCYVS	437
Qy	199	GCTGCACCTCGCTGGCTCCGCCCTCCCTGCTGCCCGAGCCCCTCTCTCGGCGCAG	258
Dd	436	YYSRYYSYSYSVSWSYSYTDYCSYRCCCYYSYSSYSYSSYSSYSSSSSSS	377
Qy	259	GAGATTCTCCCTGTGTCAGCCACCAATGGCTCTATCTCAGGAGCTGGACACTCCANG	318
Dd	376	YYVTSTNYC.T.CC...T..MCAABCSITTTTTTTT.HSCC.SA.A.M..YC.A.S	317
Qy	319	GATGGGACTAGCCCTCCAGATCCAGTCCAGTCCCTCTGGCTCCAGAATGAAGTGCCA	378
Dd	316	YSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y.YY.YYSYCSRKTW..	257
Qy	379	CCCAGCGCTGATCCAGTCTCTT	401
Dd	256	.TWTDMD.T.T..MHMY.KYB.H	234

RESULT 15

US-10-140-472-10/c

; Sequence 10, Application US/10140472

; Publication No. US20030138888A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge,Laura

; APPLICANT: Desnoyers,Jac

; APPLICANT: Filvaroff,Ellen

; APPLICANT: Gao,Wei-Qiang

; APPLICANT: Gerritsen,Mary E.

; APPLICANT: Goddard,Audrey

; APPLICANT: Godowski,Paul J.

; APPLICANT: Gurney,Austin L.

; APPLICANT: Sherwood,Steven

; APPLICANT: Smith,Victoria

; APPLICANT: Stewart,Timothy A.

```
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Applcication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-472-10

Query Match          7.6%; Score 44.8; DB 15; Length 594;
Best Local Similarity 8.7%; Pred. NO. 0.0044;
Matches 28; Conservative 135; Mismatches 160; Indels 0; Gaps 0;

QY 79 CAGAGCTACCAGCAGCCTGCTCCGCATCTCCCTAGACAAAGTCCAGCGCAGCTGGGC 138
Db 556 MM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYS.S.S.SDSYSYA.SYSYS.S.SWS 497
QY 139 CCCGAGCAGCCAGCCTCGCAGGATGCTCTCATCCATAACACCTCCAAACAGCTGCAG 198
Db 496 SYSYSSDDY.CYCCYRHHSDYSYSYY.CRCCTT.SYSRYDCHISCCSDYYCYS 437
QY 199 GCTGCACTTCGCTGCTCCGCCCTGCTCCGCCGCCGCCGCCGCCCTCTCTCTGGCGAG 258
Db 436 YSYRYYSYSYSWSYSYSDYDYSYRCCYYSYSSYSYSSAYSSTSSSSSSSS 377
QY 259 GAGGATTTCTCCTGTCAGCCACCATGCTGCTATCTCTCAGGAGCTGGACACCTCCATG 318
Db 376 YYYTSTNYC.T.CC...T.MCAABCSTTTTTTTTTT.HSCC.SA.A.M.YC.A.S 317
QY 319 GATGGAGCTGAGCCCTCAGAACTCAGTACTCCCTTGCCCTCCAGAAATGAAGTGCCA 378
Db 316 YSYS.S.S.SYMR.HRA.SHYTRS..S.MYCY.YM.Y.Y.YYSYCSKRTW... 257
QY 379 CCCAGCCTGATCCAGTCTTCTT 401
Db 256 ..TWTD.M.T.T..MHMY.KYB.H 234
```

Search completed: April 26, 2005, 11:38:08
Job time : 505 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2005, 12:22:40 ; Search time 175 Seconds
(without alignments)
573.529 Million cell updates/sec

Title: US-10-069-386A-2

Perfect score: 1033

Sequence: 1 MEGGLKRXHSDLEEEERWE.....APGSWEWELDHIMEIILGS 196

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt.03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1033	100.0	196	2	Q9UJW9	Q9UJW9 homo sapien
2	1026	99.3	196	2	Q96CQ2	Q96CQ2 homo sapien
3	871	84.3	197	2	Q9ERC3	Q9ERC3 mus musculus
4	217	21.0	236	1	STD1_MOUSE	Q9J110 mus musculus
5	213	20.6	236	1	STD1_HUMAN	Q9UHV2 homo sapien
6	208	20.1	236	2	Q6P771	Q6P771 rattus norv
7	166	16.1	244	2	Q9DCZ2	Q9DCZ2 mus musculus
8	140.5	13.6	309	1	STD2_MOUSE	Q9JJG5 mus musculus
9	136.5	13.2	248	2	Q6GM81	Q6GM81 xenopus lae
10	136	13.2	237	1	CCA4_MOUSE	Q9CWM2 mus musculus
11	133	12.9	361	2	Q6NXL9	Q6NXL9 brachydanio
12	133	12.9	383	2	Q7ZZ27	Q7ZZ27 brachydanio
13	125	12.1	314	1	STD2_HUMAN	Q14140 homo sapien
14	115	11.1	1194	2	Q76M68	Q76M68 rattus norv
15	113.5	11.0	911	2	Q80TJ8	Q80TJ8 mus musculus
16	112.5	10.9	246	2	Q6DEF1	Q6DEF1 xenopus lae
17	105.5	10.2	620	2	Q9BQ18	Q9BQ18 homo sapien
18	105.5	10.2	760	2	Q76N32	Q76N32 homo sapien
19	105	10.2	241	1	CCA4_HUMAN	Q9BXL8 homo sapien
20	104.5	10.1	740	2	Q9UPF2	Q9UPF2 homo sapien
21	101.5	9.8	728	1	P85B_HUMAN	O00459 homo sapien
22	100	9.7	846	2	Q6NVQ9	Q6NVQ9 brachydanio
23	98.5	9.5	314	2	Q6NXL2	Q6NXL2 mus musculus
24	98	9.5	517	2	Q9EEF8	Q9EEF8 anticarsia
25	98	9.5	562	2	Q8NAF0	Q8NAF0 homo sapien
26	97.5	9.4	309	2	Q8UZB4	Q8UZB4 grapevine f
27	97	9.4	803	2	Q74402	Q74402 schizosacch
28	97	9.4	1201	2	Q9C6S1	Q9C6S1 arabidopsis
29	96	9.3	817	1	VRP1_YEAST	P37370 saccharomyc
30	96	9.3	817	2	Q07229	Q07229 saccharomyc
31	95.5	9.2	432	2	Q8BN29	Q8BN29 mus musculus

ALIGNMENTS

RESULT 1

Q9UJW9
ID Q9UJW9 PRELIMINARY; PRT; 196 AA.
AC Q9UJW9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE RPA-binding trans-activator.
GN Name=RBT1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20440390; PubMed=10982866; DOI=10.1093/nar/28.18.3478;
RA Cho J.M., Song D.J., Alaoui-Jamali M.A.;
RT "RBT1, a novel transcriptional co-activator, binds the second subunit
of Replication Protein A";
RL Nucleic Acids Res. 28:3478-3485 (2000).
DR EMBL; AF192529; AAF05761.1; -
DR InterPro; IPR009263; SERTA.
DR Pfam; PF06031; SERTA; 1.
SQ SEQUENCE 196 AA; 21798 MW; 33433EF7F8A9EBA7 CRC64;

Query Match Similarity 100.0%; Score 1033; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.7e-71;
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEGGLKRXHSDLEEEERWEWSAPAGLQSYQQALLRISLDKQVRSIGPRAPSLRRHVLHN 60
Db 1 MEGGLKRXHSDLEEEERWEWSAPAGLQSYQQALLRISLDKQVRSIGPRAPSLRRHVLHN 60
QY 61 TLQQLAALRALAPALPPEPLFGEEDFSLSATIGSLTRELDTSMGTEPPQNPTPLG 120
Db 61 TLQQLAALRALAPALPPEPLFGEEDFSLSATIGSLTRELDTSMGTEPPQNPTPLG 120
QY 121 LQNEVPPQDPVFLEALSSRYLGSGLDDFFLDITDSAVEKEPARAPPEPPHNLFCAPGS 180
Db 121 LQNEVPPQDPVFLEALSSRYLGSGLDDFFLDITDSAVEKEPARAPPEPPHNLFCAPGS 180
QY 181 WEWNELDHIMEIILGS 196
Db 181 WEWNELDHIMEIILGS 196
RESULT 2
Q96CQ2
ID Q96CQ2 PRELIMINARY; PRT; 196 AA.
AC Q96CQ2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
DE RPA-binding trans-activator.

GN Name=SERTAD3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary, and Uterus;
RX MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014061; AAH14061.1; -;
DR EMBL; BC050643; AAH50643.1; -;
DR InterPro; IPR009263; SERTA.
DR Pfam; PF06031; SERTA; 1.
SQ SEQUENCE 196 AA; 21768 MW; 9C54FAA126F17AF1 CRC64;

Query Match 99.3%; Score 1026; DB 2; Length 196;
Best Local Similarity 99.5%; Pred. No. 5.9e-71;
Matches 195; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEGGLKRRKHSDEEEERWSPAGLQSQYQALLRISLDKQVRSIGPRAPSLRRHVLHN 60
DB 1 MVGGLKRRKHSDEEEERWSPAGLQSQYQALLRISLDKQVRSIGPRAPSLRRHVLHN 60

QY 61 TLQQLQAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMDGTEPPQNPTPLG 120
DB 61 TLQQLQAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMDGTEPPQNPTPLG 120

QY 121 LQNEVPPQDPVFLEALSSRYLGDGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAFGS 180
DB 121 LQNEVPPQDPVFLEALSSRYLGDGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAFGS 180

QY 181 WEWNELDHIMEIILGS 196
DB 181 WEWNELDHIMEIILGS 196

RESULT 3
Q9ERC3 PRELIMINARY; PRT; 197 AA.
AC Q9ERC3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)

DE Replication protein-binding trans-activator RBT1 (SERTA domain
DE containing 3).
GN Name=Sertad3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=C57BL/6J; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317202; AAG30951.1; -;
DR EMBL; BC034886; AAH34886.1; -;
DR MGD; MGI:2180697; Sertad3.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR009263; SERTA.
DR Pfam; PF06031; SERTA; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN 1.
SQ SEQUENCE 197 AA; 21934 MW; CFF6FA4C0A91D7E7 CRC64;

Query Match 84.3%; Score 871; DB 2; Length 197;
Best Local Similarity 85.7%; Pred. No. 4.5e-59;
Matches 168; Conservative 8; Mismatches 18; Indels 2; Gaps 1;

QY 3 GGLKRRKHSDEEEERWSPAGLQSQYQALLRISLDKQVRSIGPRAPSLRRHVLHN 60
DB 2 GGLKRRKHSDEEEERWSPAGLQSQYQALLRISLDKQVRSIGPRAPSLRRHVLHN 61

QY 61 TLQQLQAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMDGTEPPQNPTPLG 120
DB 62 TLQQLQAALRLAPALPPEPLFLGEEDFSLSATIGSLRELDTSMDGTEPPQNPTPLG 121

QY 121 LQNEVPPQDPVFLEALSSRYLGDGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAFGS 180
DB 122 PONEIVSOADPVFLEALSSRYLGDGLDDFFLDIDTSAVEKVALPPEPHNLFCAFGS 181

QY 181 WEWNELDHIMEIILGS 196
DB 182 WEWNELDHIMEIILGS 197

RESULT 4
STD1_MOUSE STANDARD; PRT; 236 AA.
ID STD1_MOUSE

RT "Generation and initial analysis of more than 15,000 full-length human
 RL and mouse cDNA sequences.";
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RP [2]
 RC TISSUE=Prostate;
 RA Struhsberg R.;
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC061808; AAH61808.1; -;
 DR InterPro; IPR009263; SERTA.
 DR Pfam; PF06031; SERTA; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 236 AA; 25419 MW; E8B7502E0473ED8B CRC64;
 Query Match 20.1%; Score 208; DB 2; Length 236;
 Best Local Similarity 31.9%; Pred. No. 3.7e-08;
 Matches 76; Conservative 33; Mismatches 67; Indels 62; Gaps 13;
 QY 4 GLKRGHSLDEEERWE-----W-----SPAGLSYQQAALLRISLDKQVRSIG 46
 DB 5 GLKRGHSLDEEERWE-----W-----SSLFDSLVVGLHSLR 58
 QY 47 PRAPSLRRHVLHNTLQQAALRLAPALPPEPL-----FLGEEDFSLSATIG 96
 DB 59 QSEPLRLHLVNVTLRIQASME--PTTVLPPELQPTAPSVADNFLSSDAGLSASWA 116
 QY 97 SILRELTSMGTEPPQNPVTLGLQNEVPP-----QPDVPFLEAL--SSRYLGDGSL 147
 DB 117 SLLEDLSHIEDLNQVQPQA-----DEGPPGKRSVGGVLPNLGALDGLGPAATGCLLDGL 170
 QY 148 DDFPDLIDTSAVEKE---PARAPPPE-PHNLFCAPGSWEW-----NELDHIMEILGS 196
 DB 171 EGLPDLIDTSMYDSELWLTPASEGLKVPEN---GPAKEETPELDEALDYLMDVLVGT 225
 RESULT 7
 Q9DC22 PRELIMINARY; PRT; 244 AA.
 AC Q9DC22;
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Mus musculus adult male kidney cDNA, RIKEN full-length enriched
 DE library, clone:061008D10 product:TRIP-BRI, full insert sequence.
 GN Name=Sertadi;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 EX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RL "High-efficiency full-length cDNA cloning.";
 RT Meth. Enzymol. 303:19-44 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 EX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RL "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690 (2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630 (2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wakahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771 (2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito K., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Segabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK002324; BAB22013.1; -;
 DR MGD; MGI:1913438; Sertadi.
 DR InterPro; IPR009263; SERTA.
 DR Pfam; PF06031; SERTA; 1.
 SQ SEQUENCE 244 AA; 26068 MW; 0F464F6419D362A6 CRC64;
 Query Match 16.1%; Score 166; DB 2; Length 244;
 Best Local Similarity 25.9%; Pred. No. 6.4e-05;
 Matches 67; Conservative 33; Mismatches 55; Indels 104; Gaps 13;
 QY 4 GLKRGHSLDEEERWE-----W-----SPAGLSQ-----YQALLRLSLDKVQVRSIGPR 48
 DB 5 GLKRN-----GREETMEALSVDSCWLDPSHPVAQTPPTVASSSLFSLVKKLHSLRQS 60
 QY 49 APSLRHRLHNTLQQAALRLAPALPPEPL-----FLGEEDFSLSATIGSI 98
 DB 61 EPDLRLHLVNVTLRIQASME--PAPVLPPELQPTAPSVADSLSSDAGLSASMASL 118
 QY 99 LRELTSMGTEPPQ-----NPVTP-----LG-----LQNEV 125
 DB 119 LEDLNHIEDLNQAPQAPQADEGPPGKRSIGGSPNLGALDGLGPAATGCLRTSI 178
 QY 126 PP-----QPDVPFLEALSSRYLGDGSLDDFDIDTSAVEKEPARAPPPEPHLPCA 177
 DB 179 PCTTVNYGYQP-----LRVSSAAPENGPAPK--BEPP----- 208
 QY 178 PGSEWNELDHIMEILGS 196
 DB 209 --ELDEALTDLMDVLVGT 225
 RESULT 8
 ID STD2 MOUSE STANDARD; PRT; 309 AA.
 AC Q9JUG5; Q8C609; Q91WL3; Q925E5;
 DT 10-OCT-2003 (rel. 42, Created)
 DT 10-OCT-2003 (rel. 42, Last sequence update)

DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
 DE MGC82098 protein.
 GN Name=MGC82098;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datschenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Krywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC074195; AAH74195.1; -;
 DR InterPro: IPR009263; SERTA.
 DR Pfam: PF06031; SERTA; 1.
 SQ SEQUENCE 248 AA; 27319 MW; E703D7E9105DF08C CRC64;
 Query Match 13.2%; Score 136.5; DB 2; Length 248;
 Best Local Similarity 30.9%; Pred. No. 0.012;
 Matches 58; Conservative 25; Mismatches 58; Indels 47; Gaps 8;
 QY 17 ERWESPAQLOYSQALLRISLKVQSRSLRPRAPSLRRHVLHNTLQQLAALRL---AP 73
 DB 10 ERECCSIPAQSH--CLANNISLVKLHRSRLRHVEPLRHFLVANTLRRLQNLQVQCAP 67
 QY 74 -----APALPEPLFLGEEDFSLSATGSLRLRDLTSMQTEPPQ 113
 DB 68 DMWKTSEBCTRSALVPESKPKPALENTEDPLSSMDASLYSSITLLEDLN-NFEGLS--S 124
 QY 114 NPVTPLGLQNEVPPQDPVFLEA-----LSSRY-----LQSGLDLDFDLID 155
 DB 125 SPLPOIEDQDLCAKPNVPVSGAEDMVKLASSSSLLSSPYLLGNLGN-LIEDIFED 183
 QY 156 TSAVEKEP 163
 DB 184 TSMYSDP 191

RESULT 10

CCA4_MOUSE
 ID CC4_MOUSE STANDARD; PRT; 237 AA.
 AC Q9CWM2; Q921B8; Q9NMP6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Cell division cycle associated protein 4 (Hematopoietic progenitor protein).
 DE NCBI_TaxID=10090;
 GN Name=Cdc4; Synonym=HEPPP;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Bone marrow;
 RX MEDLINE=22354683; PubMed=11482882; DOI=10.1006/bcmd.2001.0434;
 RA Abdullah J.M., Jing X., Spassov D.S., Nachtman R.G., Jurecic R.;
 RT "Cloning and characterization of Hepp, a novel gene expressed preferentially in hematopoietic progenitors and mature blood cells.";
 RL Blood Cells Mol. Dis. 27:667-676(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Meioesophros;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reid J., Ring B.Z., Ringwald M.,
 RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyszewski-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [3]
 RP SEQUENCE OF 13-237 FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datschenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC074195; AAH74195.1; -;
 DR InterPro: IPR009263; SERTA.
 DR Pfam: PF06031; SERTA; 1.
 SQ SEQUENCE 248 AA; 27319 MW; E703D7E9105DF08C CRC64;

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: May be involved in molecular regulation of hematopoietic
 CC stem cells and progenitor cell lineage commitment and
 CC differentiation.
 CC -!- TISSUE SPECIFICITY: Expressed preferentially in hematopoietic
 CC progenitors and mature blood cells. Expressed at low levels in the
 CC heart, lung, spleen, and thymus and at a higher level in muscle.
 CC -!- DEVELOPMENTAL STAGE: Developmentally regulated. Preferential
 CC expression in both fetal and adult hematopoietic progenitors and
 CC mature blood cells during embryonic and adult hematopoiesis.
 CC -!- SIMILARITY: Belongs to the TRIP-Br family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AK010535; BAB27012.2; -;
 DR EMBL; AF322238; AAK31074.1; -;
 DR EMBL; AK032980; BAC28109.1; -;
 DR EMBL; BC012953; AAH12953.1; ALT_INIT.
 DR MGD; MGI:191213; Cdc44.
 DR InterPro; IPR009263; SERTA.
 DR Pfam; PF06031; SERTA; 1.
 DR CONFLICT 196 196 S -> T (in Ref. 3).
 FT CONFLICT 213 213 T -> A (in Ref. 3).
 FT CONFLICT 213 213
 SQ SEQUENCE 237 AA; 26107 MW; 52EC046EEDDF326E CRC64;
 Query Match 13.2%; Score 136; DB 1; Length 237;
 Best Local Similarity 26.1%; Pred. No. 0.012;
 Matches 63; Conservative 32; Mismatches 86; Indels 60; Gaps 9;
 QY 4 GLKRGHSDLEEEEREWSPA-GLQSYOQALLRLSLDKVQSLGPRAPSLRRVLIHNTL 62
 DB 5 GLKRGYDOEEGVEGFGTVPSYLSQ--RQSLDMSLVKLQCHMLVFNLCRSVLIAITV 62
 QY 63 QOLQAALR-----LAP--APALPPEPLFLGEEDFSLSATIGSLIRELDTSMGTEPPQ 113
 DB 63 RQIQEEMSQDQVHWGMAQNVDRAPVERLVSTE---ILCRTVRGAEERHAPAELEDAPLQ 119
 QY 114 NPVTPLGLQNEVPPQDP-----VFLEALSSRYLGDGLDFF 151
 DB 120 NSVSELPVGSAPGQNPQSSILWEMDSPQENRGSFQKSLDQIFETLENK--NSSSVSELF 177
 QY 152 LDIDTSAYEKEPAR-----APPEPPHNLFCAPGSWEWNELDHIMEII 193
 DB 178 SDVSSSYDLPTVLTMGMSGTKSLCNGLEGFAAATPPPSSTCKS---DLAELDHVVEIL 234
 QY 194 L 194
 DB 235 V 235
 RESULT 11
 Q6NXd9 PRELIMINARY; PRT; 361 AA.
 AC Q6NXd9;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein sertad2.
 GN Name=sertad2;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC067135; AAH67135.1; -;
 DR ZFIN; ZDB-GENE-030616-569; sertad2.
 DR InterPro; IPR009263; SERTA.
 DR Pfam; PF06031; SERTA; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;
 Query Match 12.9%; Score 133; DB 2; Length 361;
 Best Local Similarity 22.3%; Pred. No. 0.035;
 Matches 81; Conservative 35; Mismatches 71; Indels 176; Gaps 17;
 QY 4 GLKRGHSDLEEEEREW-SPAGLQ-----SY---QOALLRLSLDKVQSLGPRAPS 51
 DB 5 GAKRK--LDDESGLEGKALAGAGAGGLSKVSYTLQRTIFNMSLMKLYNHRVTEPS 61
 QY 52 LRRVLIHNTLQQQAALR-----PEPL-----FLGEEDFSLSA-----L 71
 DB 62 LEKVLINNNLRRITQDELKQGNLRPLFFPPSPPPDPVDESFPREPQPAFVLSWVAPPI 121
 QY 72 APALP-----PEPL-----FLGEEDFSLSA----- 93
 DB 122 SQSALGASSLTPSSGLSNPAFLACITPAPLLEEDNVSLCTSPSLAPPAPTSRLSPS 181
 QY 94 ----TIGSILRELD-----TSMGTEPP-----QNPVTPL 119
 DB 182 VARDSSALDEIEELCPSPLTATSPAGATSPSPQLCPPLNSGALDSDCKSPK 241
 QY 120 --GL-----QNEVPPQDP---PVFLE-----ALSSRYLGDGLD--FFLDIDTSAYEK 161
 DB 242 LEGLVPLAERSAVENTPETLPNLSLDMSSTSPSSSGFLTDALDDILFDIDTSMYDF 301
 QY 162 EP-----ARAPPEPPHNLFCAPGSWEWNELDHIMEII 193
 DB 302 DPTSSSGAAPSKLAPWTADELKLTSPSPYSGAAPVSSN---QPFKMDLTDLHIMEVL 358
 QY 194 LGS 196
 DB 359 VGS 361
 RESULT 12
 Q7ZZ27 PRELIMINARY; PRT; 383 AA.
 ID Q7ZZ27


```
AC Q7Z227;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE SI:zC10IN13.8 (Novel protein similar to human solute carrier family 1
DE (Neuronal/epithelial high affinity glutamate transporter, system Xag),
DE member 1 (SLC1A1)).
GN Name=sertad2; Synonyms=SI:zC10IN13.8;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Kimberley A.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL772154; CAD61094.1; -.
DR ZFIN; ZDB-GENE-030616-569; sertad2.
DR InterPro; IPR009263; SERTA.
DR Pfam; PF06031; SERTA; 1.
SQ SEQUENCE 383 AA; 40402 MW; 76514201CDFC992F CRC64;

Query Match 12.9%; Score 133; DB 2; Length 383;
Best Local Similarity 22.3%; Pred. No. 0.037;
Matches 81; Conservative 35; Mismatches 71; Indels 176; Gaps 17;

QY 4 GLKRRKHSDLLEEEERWEMW-----SPAGLQ-----SY---QQALLRISLDKQVRSGLGPRAPS 51
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
27 GAKRK--LDEDEGLEGKALAGAGAGLSKVSYTLQRTIFNMLKLYNHRVATPS 83
QY 52 LRRHVLHNTLQOALR-----PPPL-----FLGEDEFSLSA-----L 71
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
84 LKRVLLNNLMRRIODELQEGNLRPLFPFPPPPDDPDVDESFRFPQAFSVLSMWAPPI 143
QY 72 APAPALP-----PPPL-----FLGEDEFSLSA-----93
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
144 SOSPALSASLTSPPSSGLSNAPLACLTAPSLLEDNVSLCTSPLAPAPTSLRLSPS 203
QY 94 ----TIGSLRLD-----TSMGTEPP-----QNPVTPL 119
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
204 VARDSPSSALDIIEELCPSPLTATSAFCATSPPLQCPPLNSGALDSKCSKPCPK 263
QY 120 --GL-----QNEVPQPD---PVFL-----ALSSRYLGDGLDD--FFLDIDTSAVEK 161
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
264 LEGLVPLAERSAVENTPETLPNSLDMSTSPSSSGFLTDALDILFADIDTSMYDF 323
QY 162 EP-----ARAPPEPHNLFCAPCSWENNELDHIMEII 193
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
324 DPTCTSSGAAPSKLAPMTVADLLKTFSPYSGAAPVSSN---QPPKMDLTLDHIMEVYL 380

QY 194 LGS 196
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
381 VGS 383

RESULT 13
STD2 HUMAN STANDARD; PRT; 314 AA.
AC Q14140;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE SERTA domain-containing protein 2 (Transcriptional regulator
DE interacting with the PHD-bromodomain 2) (TRIP-Br2).
GN Name=sertad2; Synonyms=KIAA0127;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;

RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
CC -!- FUNCTION: Acts at E2F-responsive promoters to integrate signals
CC provided by PHD- and/or bromodomain-containing transcription
CC factors (by similarity).
CC -!- SIMILARITY: Belongs to the TRIP-Br family.
CC -!- SIMILARITY: Contains 1 SERTA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D50917; BAA09476.2; ALT_INIT.
DR Genew; HGNC:30784; SERTAD2.
DR InterPro; IPR009263; SERTA.
DR Pfam; PF06031; SERTA; 1.
KW Transcription regulation.
SQ SEQUENCE 314 AA; 33896 MW; E43107FC565AAC31 CRC64;

Query Match 12.1%; Score 125; DB 1; Length 314;
Best Local Similarity 22.7%; Pred. No. 0.12;
Matches 71; Conservative 34; Mismatches 88; Indels 120; Gaps 12;

QY 2 EGGLKRRKHSDLLEEEERWEMWSPAGLQS-----YQALLRISLDKQVRSGLGPRAPSLRRH 55
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4 KGG-KRKFDDEHDEGLGKIVSPCDGSPKSVYTLQRTIFNMLKLYNHRPLTEPSLQKT 62
QY 56 VLIHNTLQOALR-----70
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
63 VLIHNTLQOALR-----70
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
71 ----LAPA-----PALPPE---PLFLGEDEFSLSATIGSLRLD 103
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
123 TTPLEACLTAPSLLEDLDDDDTCTSQAMQPTATKLSPALLPEKD-SFSSALDEIELCP 181
QY 104 TSMD-----GTEPPQNVPVTLQNEVPQ---PDVFLRAL-----SSRYLGD 145
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
182 TSTTEAATAATDSVKGTSSAGTQKLDGQPSRADDSKLDLPGNFFITTTGFLTDL 241
QY 146 GLDD--FFLDIDTSAVEKEP-----ARAPPEPHNLF-----CAPCSW 183
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
242 TLDDILFADIDTSMYDFDPTCTSSSGTASGMAPVSADLLKTLAPYSSQPVTPSPQPMDL 301

QY 184 NELDHIMEIILGS 196
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
302 TELDHIMEIILGS 314

RESULT 14
Q76M68 PRELIMINARY; PRT; 1194 AA.
AC Q76M68;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE SynArfGEF.
GN Name=sag;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Inaba Y., Zhang J.P., Tian Q., Okano A., Li W., Komiyama A.,
RA Suzuki T.;
```

RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AB057643; BAD14305.1; --
 DR HSSP; Q15438; 1BC9.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR000904; Sec7.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF01369; Sec7; 1.
 DR SMART; SM00222; Sec7; 1.
 DR PROSITE; PS00096; IQ; 1.
 DR PROSITE; PS0190; SEC7; 1.
 SQ SEQUENCE 1194 AA; 129017 MW; 4828FB1AA6F78B56 CRC64;

Query Match

Best Local Similarity 11.1%; Score 115; DB 2; Length 1194;
 Matches 58; Conservative 19; Mismatches 64; Indels 98; Gaps 11;

QY 5 LKRKHSLEE-EEERWWS-----PAGLSYQQALLRLSLDKVORSIGPRAP 50
 DB 976 LKESIAEVTLEQIRIEWELERQOGTKTLSARSAGAGDPQSKQGSPTAKREAMAGEKAT 1035

QY 51 SLRRHVLIHNTLQQLQAALRL-----APAPALPPEPLFLGDEDFSLSATIGS 97
 DB 1036 ESSGEVSIHNRLOTQHSKPGVGERGAPAPSPPTSPPPPLPPDPQ-----PS 1082

QY 98 ILRELDTSMDGTPEPPQNPVTPLGLQNEVPP-----QDPVFLFA 136
 DB 1083 PLRE-----QPP-----PLPLPPTPPGTIVQCCQIVKIVLDKPCCLARMEPLLSQA 1129

QY 137 LSSRYLGSGLDDFFLDIDTSA-----VEKEPARAPPEPPHN---LFCAPGS 180
 DB 1130 LSC-YASSSS-----DSCGSTPLRGPGSPVKVIHQPLPPPPPPYNNHPHQFCPPGS 1179

RESULT 15

Q80TJ8 PRELIMINARY; PRT; 911 AA.
 AC Q80TJ8;
 DT 01-JUN-2003 (TReMBLrel. 24, Created)
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE MKIAA110 protein (fragment).
 GN Name=AK122446; Synonyms=mkIAA1110;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 RA Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT cDNAs identified by screening of 400 mouse KIAA-homologous
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:35-48(2003).
 DR EMBL; AK122446; BAC65728.1; --
 DR HSSP; Q15438; 1BC9.
 DR MGD; MGI:2677208; AKJ22446.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000904; Sec7.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF01369; Sec7; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00222; Sec7; 1.
 DR PROSITE; PS00096; IQ; 1.
 DR PROSITE; PS0190; SEC7; 1.
 FT NON_TER 1
 SQ SEQUENCE 911 AA; 99413 MW; 8D267BC6C46382A6 CRC64;

Query Match

11.0%; Score 113.5; DB 2; Length 911;

Best Local Similarity 25.2%; Pred. No. 3.4;
 Matches 60; Conservative 19; Mismatches 62; Indels 97; Gaps 12;
 QY 5 LKRKHSLEE-EEERWWS-----PAGLSYQQALLRLSLDKVORSIGPRAP 50
 DB 694 LKESIAEVTLEQIRIEWELERQOGTKTLSARSAGAGDPQSKQGSPTAKREAMAGEKAA 753
 QY 51 SLRRHVLIHNTLQQLQAALRL-----APAPA-----LPPEPLFLGDEDFSLSATIGSI 98
 DB 754 ESSGEVSIHNRLOTQHSKPGVGERGAPAPSPPTSPPPPLPPDPQ-----PSP 800
 QY 99 ILRELDTSMDGTPEPPQNPVTPLGLQNEVPP-----QDPVFLFA 137
 DB 801 LRE-----QPP-----PLPLPPTPPGTIVQCCQIVKIVLDKPCCLARMEPLLSQA 847
 QY 138 SRYLGSGLDDFFLDIDTSA-----VEKEPARAPPEPPHN---LFCAPGS 180
 DB 848 SC-YASSSS-----DSCGSTPLRGPGSPVKVIHQPLPPPPPPYNNHPHQFCPPGS 896

Search completed: April 25, 2005, 12:32:29
 Job time : 178 secs